ISOLATION OF HIGH AFFINITY LIGANDS AGAINST SH3 AND EVH1 DOMAINS USING THE PHAGE DISPLAY COSMIX-PLEXING® TECHNOLOGY

Von der Gemeinsamen Naturwissenschaftlichen Fakultät
der Technischen Universität Carolo-Wilhelmina
zu Braunschweig
zur Erlangung des Grades einer
Doktorin der Naturwissenschaften
(Dr. rer. nat.)
genehmigte
D i s s e r t a t i o n

von Nathalie HORN (geb. Chavand)
aus Grenoble (Frankreich)

1. Referent: Prof. Dr. John Collins
2. Referent: Prof. Dr. Rüdiger Cerff
eingereicht am: 16.03.2000
mündliche Prüfung (Disputation) am: 06.06.2000
Druckjahr: 2000

A mes parents
A Clemens

TABLE OF CONTENTS

FIGURE LIST

I - Introduction

III - Results

TABLE LIST

I - INTRODUCTION

A – Combinatorial libraries for affinity selection of novel ligands

Protein-ligand interactions form the basis of almost all cellular functions. All processes, in living organisms, are tightly controlled by highly specific interactions between biological macromolecules. The identification of disease relevant target molecules, and the selection and optimization of specific ligands, is therefore a focus of much of current biochemical research, and a prerequisite for most pharmaceutical applications. Various approaches have been developed during recent years, to improve our understanding of the principles governing molecular interactions.

A conventional method, the so-called ”rational design”, is performed by computer modeling of molecular structure, prediction of three dimensional alterations which may follow on individual amino acid exchanges. This is followed by construction of the corresponding DNA and expression of the recombinant protein for testing and evaluation. This approach seeks to design proteins for specific tasks by drawing correlations between amino acid sequence and specific folding in an attempt to understand a complex system, through its topography. However, this technique is limited by several factors. Among them, the knowledge about the structure/function of most proteins or molecules within the cell is not sufficient. There is a poor correlation between *in vitro* results, and the computer predictions, due to the complexity of parameters involved. Another main difficulty is the astronomical number of relevant mutants which might be designed, in order to optimize a given function (Pisabarro and Serrano, 1996).

In contrast to rational design procedures, nature employs mutations, and selections to evolve highly adapted individuals. Recent technological advances have demonstrated that it is possible to mimic these natural evolutionary processes. Current trends in the search for novel pharmaceutical agents have focused on generating combinatorial libraries as large as 10^{12} molecules, using *in vitro* evolution techniques. Combinatorial libraries are a very powerful tool to identify highly specific and affine binders to particular targets. Repeated rounds of screening and amplification of candidates, have produced novel molecules capable of binding their targets with higher affinity than their natural counterparts. The main limitation of these libraries, is their low complexity. The greater the diversity of a bank, the more likely it will contain a ligand with optimized properties to interact with a given molecule. Therefore, increasing numbers of researchers devote their work to enhance the *in vitro* evolution

processes, and thus, the diversity of the library. Several approaches, which will be discussed in more details later, have been successfully developed to create mutations, such as error prone PCR, or directed mutagenesis. However, the use of recombination, with or without somatic mutagenesis, mimics further the evolutionary process in natural systems, and generates additional diversity. Using these techniques (chain shuffling, DNA shuffling, and cosmix-plexing®), one can generate reassortment between mutagenized regions and accumulate modifications over larger areas in complex proteins.

Different sorts of combinatorial libraries are in use, which can be composed either of synthetic or biosynthetic molecules. Several of these libraries are briefly described in this work. However, the most powerful of these systems, the phage display libraries, with respect to the size of the banks, the range of applications, or the possibility to reproduce the natural evolution, will be discussed in more details.

1- Combinatorial chemistry libraries

During the last few years, a variety of methodologies have been developed that permit simultaneous synthesis of multiple peptides. It is thus possible to generate the large number of different sequences required for a library. The generation of molecular diversity using strategies that covalently connect together members of a set of chemical building blocks, in all possible combinations, represents a revolution in multiple synthesis (for a review on multiple peptide synthesis, see Fields et al., 1990). The applications of these peptide libraries include investigating protein/protein (Kramer and Schneider-Mergener, 1995; Reineke et al., 1998), protein/DNA (Kramer et al., 1993; Reuter et al., 1999), and protein/metal interactions (Malin et al., 1995). In addition, substrate specificities of kinases (Toomik et al., 1996; Tegge et al., 1998; Mukhija et al., 1998), proteases (Duan and Laursen, 1994), and chaperones (Rüdiger et al., 1997) can be investigated.

The problem of finding peptide ligands that bind to monoclonal antibodies which recognize discontinuous epitopes within a protein antigen was first approached by Geysen using a combinatorial synthesis procedure referred to as the “mimotope strategy” (Geysen et al., 1986). In this strategy, peptide mixtures are synthesized in the format: $XXXO_1O_2XXX$, where positions O_1 and O_2 are defined amino acids, while the X positions are randomized. 400 sets of peptides of this form are generated, in order to ensure the presence of all possible O_1O_2 combinations within the peptides. The preferred set for a given target, is defined (for example by ELISA), and used as a starting point for the creation of the next subset of libraries, of the form $XXO_3D_1D_2O_4XX$, where D_1 and D_2 are the positions previously defined,

and which are now kept constant. O_3 and O_4 are their direct flanking positions, and are presently scanned: all possible combinations are prepared on these positions. These cycles of synthesis and screening are reiterated until the entire sequence is optimized, and the resulting peptide is termed a “mimotope” (for epitope mimetic) for the target.

This strategy has been modified by Dooley and Houghten, in order to avoid the time consuming resynthesis of the sub-libraries. For this purpose, a moderate number of very large synthetic libraries are generated, by introducing only one defined position, and coupling at all other positions, mixtures of the building blocks (O_1XXXXX; XO_1XXXX; XXO_1XXX...). The different pools are subjected for testing for the library which gives the strongest response. The optimized peptide for a particular target is obtained by aligning the defined positions of the best libraries isolated (Dooley and Houghten, 1993).

Frank et al. (1995), proposed a modified strategy of the previous one, using a larger number of pools, each consisting of a smaller component. The pools are designed in such a way that two positions are defined, whereas the other ones are randomized (O_1O_2XXXX; XO_1O_2XXX; XXO_1O_2XX...). Matching the overlaps of these dipeptide sequence preferences for a particular target, permits to conceive the optimized sequence for this target.

The different synthetic methods can adopt the preferred combinatorial strategies. One of these methods, the SPOT method, developed by Ronald Frank, presents a strong advantage in its ease to manipulate. The Spot synthesis is a flexible technique for simultaneous, parallel chemical synthesis on membrane supports (Frank, 1992). The synthesis areas (spots) are defined on a cellulose membrane, by spotting a Fmoc-β-alanine-pentafluorophenyl ester solution. Routinely, peptides up to a length of 20 residues can be synthesized with sufficient fidelity. The peptides can be used for binding assays directly on the membrane, or cleaved from the solid support, for use in solution (for a review, see Reineke et al., 1999). The analysis between the peptides immobilized on the membrane, and the target of interest, is done in a Western blot-like manner. Positions at which the target was detected can be directly used to determine the sequence of the peptides with affinity for the target.

The combinatorial chemistry libraries were shown to be successful in epitope mapping, and in defining T-cell epitopes. However, the size of the libraries are at the limits of the technology with a diversity of approximately 10^8 variant molecules per ml. Furthermore, once a library has been used to check for peptides with particular features for a given target, it is not possible to reuse it. New libraries have to be generated for further work with the target of interest. In addition, new libraries are required for each target that has to be tested. This is time consuming, and increases the costs of the analysis.

2- Yeast two hybrid system

The two hybrid system is designed to detect and study protein-protein interactions *in vivo*. The basic strategy of the two hybrid system is the modular architecture of a group of eukaryotic transcription factors, of which the yeast GAL4 transcription factor is the prototype. These factors consist of a DNA binding domain which confers specificity for a given promoter sequence, and an activation domain that interacts and activates the transcription machinery.

The DNA binding domain binds DNA in a sequence specific manner, but fails to activate transcription. The activation domain cannot activate transcription on its own, as it is not sufficient to localize and bind the DNA of interest.

The two hybrid system offers advantages to study protein-protein interactions, since it is possible to bridge the two domains through a secondary interaction, without disrupting their function (Ma and Ptashne, 1987). Two hybrid proteins, each containing one of two interacting proteins, fused with the DNA binding domain, or with the activation domain of the GAL4 transcription factor (as instance), can be created. When the two interacting proteins make a contact, the two domains of the GAL4 are brought together, and their function is restored (Fields and Song, 1989; for a review about new hybrid technologies: Kolanus, 1999). It is then possible to discriminate which hybrid interacts, using the expression of a reporter gene product as signal.

The main drawback of the two hybrid system is that this method found applications only when potential binders for a target are known. Then, the technique can be used to check for this interaction, or to search for optimized binding peptides, by fusing to one domain of the transcription factor, peptides with modified positions. Then, it is possible to screen for the altered peptide which shows the best binding characteristics for the target. It has the advantage for the analysis of intracellular peptides and proteins in that they are produced with the correct folding, in this system.

3- Ribosome display

An *in vitro* system based on the display of nascent peptides in ribosome complexes has been developed to circumvent the limitation in library diversity faced by peptide libraries. Mattheakis and coworkers obtained successful results with this technique. They constructed a molecular DNA library encoding 10^{12} decapeptides, and expressed the library in an *E.coli* S30 *in vitro* coupled transcription/translation system. Conditions were chosen to ensure that the poly-ribosomes containing nascent peptides remained attached to their encoding mRNA.

Therefore, when a complex shows good affinity for a target, the RNA can be recovered from the bound complex, and converted into cDNA. The cDNA from the enriched pool of ribosomes (after several rounds of screening), has then to be cloned into a phagemid vector. This vector serves as both a peptide expression vector, and as a DNA sequencing vector, for peptide identification. By expressing the ribosome derived peptides on phage, one can either continue the affinity selection procedure in this format, or assay the peptides on individual clones for binding activity in a phage ELISA test (Mattheakis, et al., 1994). For a recent review about ribosome display, one can refer to Hanes and Plückthun (1999).

Szostak developed a new strategy to use ribosome display, where a stable linkage between the informational (mRNA) and the functional (encoding peptide) domains of the resulting joint molecules is established (Roberts and Szostak, 1997; Nemoto et al., 1997). A covalent fusion between an mRNA and the peptide that it encodes, is generated by *in vitro* translation of synthetic mRNAs that carry puromycin, a peptidyl acceptor antibiotic, at their 3´ end. The recycling is performed via error prone PCR, which introduces further mutations. The advantage of this technique, is the large number of variants obtained. In addition, it allows a specific mRNA to be enriched from a complex mixture of mRNAs based on the properties of its encoded peptide. The drawbacks are that the initial binding kinetics are very poor, and that there is a low enrichment rate per round.

4- Bacterial surface display

Several systems have been described for the bacterial display of antibody fragments (see Georgiou et al., 1997, for a review). The cell envelope of *E.coli* and other Gram-negative bacteria consists of the inner membrane, the peptidoglycan cell wall, and the outer membrane. Although the outer membrane serves as a barrier to protein secretion, Francisco and coworkers constructed a chimeric targeting sequence that directs fused proteins to the cell surface. The chimeric targeting sequence consists of the leader peptide and first 9 amino acids of the *E.coli* major outer membrane lipoprotein, and amino acids 46-159 of the outer membrane protein OmpA (Francisco et al., 1993). The function of the lipoprotein is to direct the chimera to the outer membrane. The OmpA region transverses the membrane, bringing the fusion protein on the cell's exterior. Lipoprotein fusions (without partial OmpA) was found to be detrimental for the integrity of the cell envelope, caused extensive cell lysis, and was tethered to the interior face of the outer membrane, and thus, not exposed to the extracellular fluid. These problems were addressed by using the OmpA in correlation with the lipoprotein.

In the pAP system, Fuchs et al., expressed a scFv as fusion protein with the peptidoglycan-associated-lipoprotein (Fuchs et al., 1996). This fusion protein binds strongly to the cell wall, and is displayed by the bacteria.

The ability of the fused protein to bind to the target of interest can be determined with methods such as ELISA. However, screening of the cell surface displayed libraries is faster and more efficient when using fluorescence activated cell sorting (FACS). The libraries are incubated with the fluorescently-labeled target molecule. No elution step is required, ruling out some possible complications, including nonspecific binding to the support material, or inability to elute the very tightly binding clones. Cell sorting provides two distinct advantages. In principle, a bacterial display system might provide a high enrichment factor of positive clones, since it is possible to select one bacterium stained with a fluorescently labeled antigen, from 1000 to 10000 unstained bacteria, after only one passage through a FACS. The second benefit of the FACS, is the ability to discriminate directly between binders of different affinity and specificity. It has been shown that mean fluorescence intensity of cells displaying the fusion proteins varies proportionally to the antigen affinity for the chimeric protein.

Libraries are propagated by growing the cells in liquid cultures. Selected clones are purified and amplified in one step by plating on agar.

However, the bacterial display has severe limitations. First, the cell surfaces are very complex. Then, the size of the library is restricted to 10^5 clones, due to limitation of the transformation efficiency. In addition, only secreted peptides or proteins can be presented. Furthermore, it appears with the peptidoglycan-associated-lipoprotein system, that large proteins, along with, small ones, cannot be used as antigens. Large antigens are unsuitable due to their limited accessibility for the system, whereas small proteins bind unspecifically, probably depending on their charge or hydrophobic character. In the lipoprotein-OmpA system, restrictions occur, as the amount of surface expressed fusion protein was reported to be critical for viability. Moreover, the relatively large number of antibodies per cell (50 000 to 100 000 copies) present in both systems, might be an inconvenient. The polyvalent presentation might lead to avidity problems, and prevent the isolation of clones with high affinity for the target molecules.

B – Phage display

Phage display technology represents a system in which filamentous bacteriophage display proteins or peptides on their surface. The resulting libraries exhibit an enormous

molecular diversity, and are used as a huge source of sequences to select for molecules possessing the essential structure(s) to bind a particular target. Thus, this method is particularly convenient to understand protein-protein interactions, and structure-function relationships.

To facilitate a discussion of how the variant proteins or peptides are efficiently expressed on the bacteriophage surface, a description of the biology of the bacteriophage is presented.

1- Life cycle of bacteriophage

1-1 The Ff bacteriophage particle

Filamentous phage are virus particles able to infect a variety of gram negative bacteria, using pili as receptors. They are temperate phage living in symbiosis with their host cell, continuously producing phage particles, without lysing the host. The best characterized of these phage are M13, fl, and Fd, which infect *Escherichia coli* containing the F conjugative plasmid. Because of their similarity (Van Wezenbeek et al., 1980; Beck and Zink, 1981; Hill and Petersen, 1982), and their dependence on the F plasmid for infection, M13, fl, and Fd are collectively referred to as the Ff phage.

The Ff phage genome has 9 genes which encode for 11 proteins. The genes are grouped on the DNA according to their functions in the life cycle of the bacteriophage. Three sets are distinguished, which respectively encode for DNA replication proteins (pII, pV, and pX), capsid proteins (pIII, pVI, pVII, pVIII, and pIX), and for proteins involved in the membrane associated assembly of the phage (pI, pIV, and pXI). A short stretch of DNA, called “intergenic region”, contains the site of origin for the synthesis of the viral and complementary DNA strands (+/-), as well as a hairpin region which is the site of initiation for the assembly of the phage particles (packaging signal) (Webster, 1996).

Ff phage have a single stranded, covalently closed DNA genome (6407 nucleotides) which is encased in a long cylinder approximately 7 nm wide by 900 to 2000 nm in length. This flexible cylinder is composed of many copies of the major coat protein (pVIII ca. 2700 units), with several minor proteins (pIII, pVI, pVII, and pIX), located on the extremities. One end of the particle contains about 5 molecules each of pVII and pIX. The DNA is oriented within the virion with the packaging signal located at the pVII and pIX end of the particle. It is suggested that only pIX is exposed at the end of the particle, and pVII is buried.

At the other end of the particle, about 5 copies each of pIII and pVI are present. The

amino-terminal portion of pIII is required for the infection process. It is proposed that the hydrophobic amino terminus of pVI is buried within the particle. The observation that phage particles containing foreign proteins fused to the carboxyl end of pVI can be produced would argue that the carboxyl terminus of pVI is very near the surface of the virion (e.g. pDONG61 vectors; Jesper et al., 1995).

The pVII and pIX proteins are required for efficient particle assembly, while pVI and pIII are crucial for particle stability and phage infectivity.

1-2 The infection process

Infection of *E.coli* by the Ff phage is initiated by the specific interaction of pIII with the tip of the F conjugative pilus. pIII is sufficient to mediate the phage infection, and has, therefore, been intensively studied (Holliger and Riechmann, 1997), to help the understanding of the infective process. The pIII protein is divided into three domains separated by glycine-rich regions. It contains a short C-terminal transmembrane segment (figure I-1). The N-terminal domain (D1) is thought to be responsible for membrane penetration, the middle domain (D2) for adsorption to the F-pilus tip, and the C-terminal domain (D3) for anchoring the pIII in the phage particle.

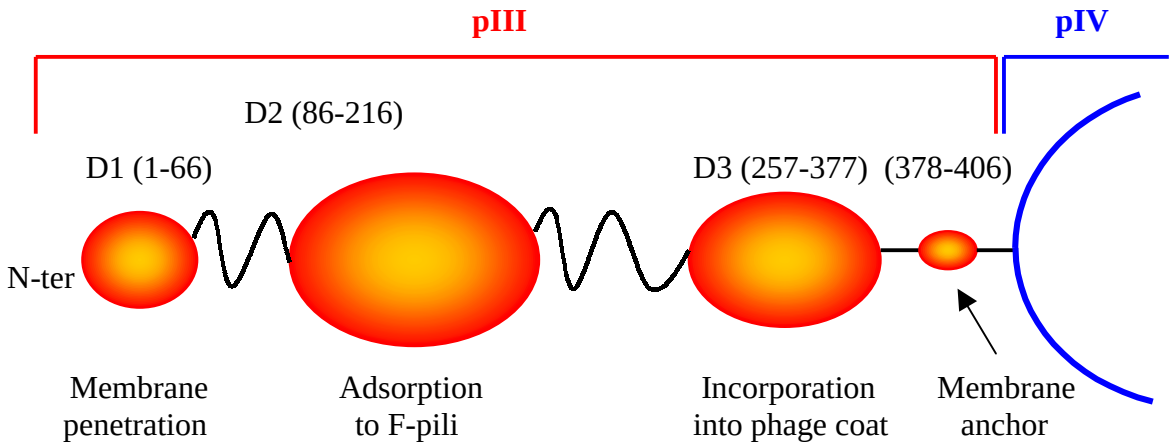


Figure I-1: The minor coat protein pIII of Ff phage. Schematic organization and sequence localization of domains. The numbers refer to the amino acid contained by each domain (Holliger and Riechmann, 1997).

The F pilus consists of a protein tube made up of pilin subunits which are assembled and disassembled by a polymerisation and depolymerisation process in the bacterial inner membrane. Pilus retraction (by depolymerisation) draws the donor and recipient bacteria together: the end of the phage attached to the F pilus, is brought to the membrane surface. The

major capsid proteins (pVIII) integrate into the inner membrane, and the phage DNA is translocated into the cytoplasm. The translocation event requires the product of the *tolQRA* genes. The tolQRA proteins are located in the inner membrane. They may form a complex which can communicate between the inner and outer membranes. One of the normal functions of these Tol proteins may be to maintain the integrity of the outer membrane, to avoid a leak of periplasmic proteins into the media.

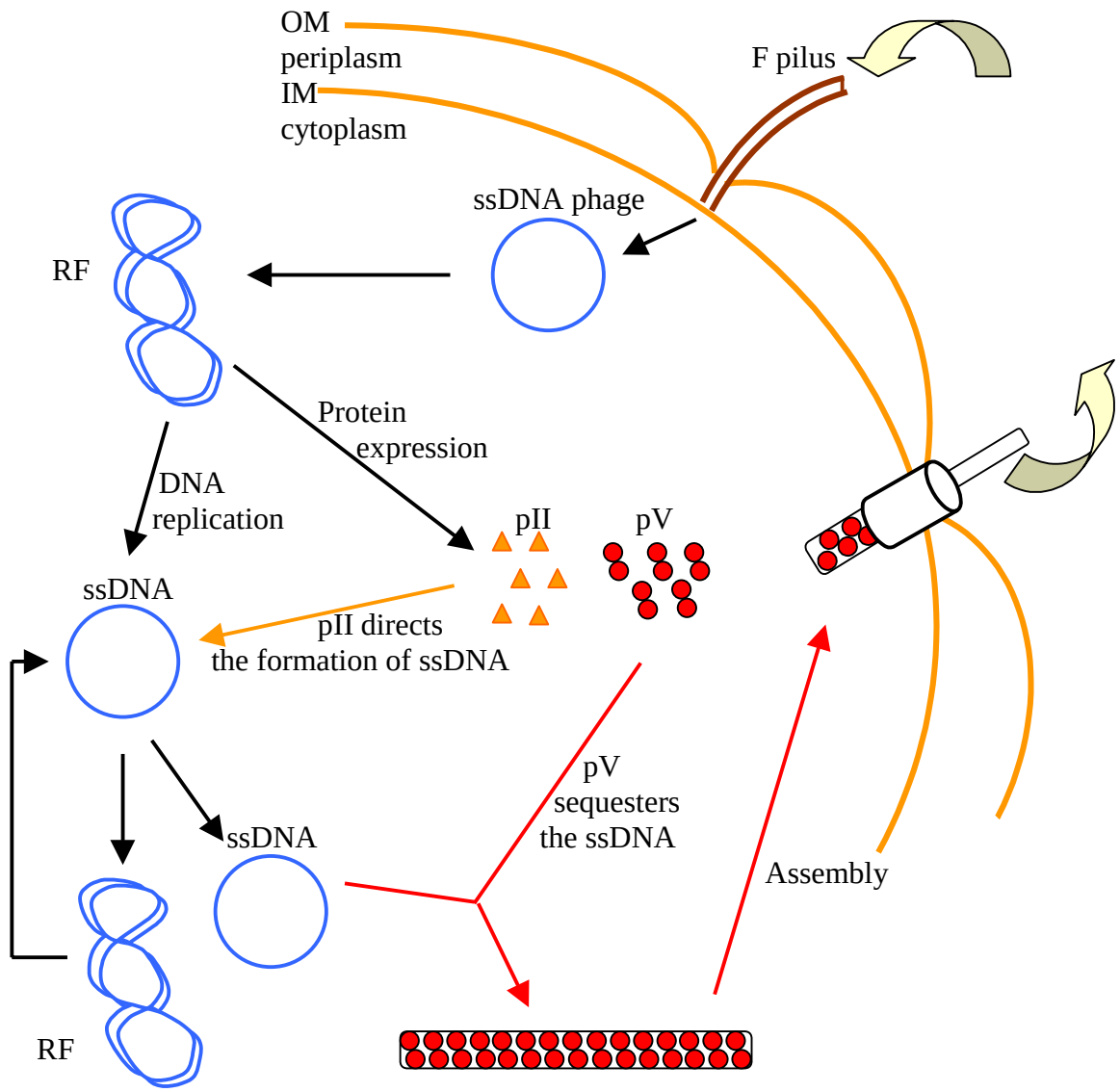


Figure I-2: Schematic representation of the life cycle of the Ff bacteriophage. The phage particle enters the host cell through the F pilus, and is translocated into the cytoplasm. Bacterial enzymes synthesize the complementary strand (replicative form: RF). The RF molecule serves as template for transcription, and translation of the phage proteins. pII and bacterial enzymes direct the synthesis of single-stranded phage DNA. The newly synthezised single-stranded DNA either is converted to another RF DNA by host enzymes, or, when pV phage protein reaches the proper concentration, is sequestered into a complex which is assembled into new phage particles. The phage are released into the media. OM and IM, outer and inner membrane; RF, supercoiled double-stranded replicative form DNA; pII and pV, products of phage genes II and V, respectively (adapted from Webster, 1996).

After the phage DNA translocation, bacterial enzymes convert the infecting single-stranded phage DNA into a supercoiled, double stranded replicative form (RF) molecule (figure I-2). pII binds and introduces a specific cleavage on the (+) strand of the RF molecule. The resulting 3' end is used by the host enzymes as a primer in order to synthesize a new (+) strand which is converted in a new RF DNA by bacterial enzymes. pX is crucial for the proper replication of the phage DNA, and functions in part as an inhibitor of pII function (Fulford and Model, 1984). All phage proteins are synthesized from the RF molecule, and their production increases with the accumulation of the RF molecules. Phage proteins involved in the assembly of the particle become integrated into the cell envelope while proteins involved in DNA replication remain in the cytoplasm.

When the pV (phage specific single-stranded DNA binding protein) reaches the proper concentration, it sequesters into a complex, the newly synthesized phage single-stranded DNA, obtained due to a specific cleavage on the (+) strand of the RF molecule, directed by pII. pV binding prevents the conversion of the newly synthesized single-stranded DNA to RF. The pV/DNA interaction give a structure of approximately 880 nm long, and 8 nm in diameter, containing one (+) strand, and around 800 pV dimers. The packaging signal is at one end of this structure, available for the initiation of phage assembly (Model and Russel, 1988).

1-3 The assembly process

The assembly process occurs at the bacterial envelope, at a site where the inner and outer membranes are in close contact. Interaction of the pV-phage DNA complex with proteins in the assembly site is the initiating event for assembly of the particle. Assembly proteins pI, pIV, and pXI are required for phage assembly, but are not present in the phage particle. pI monomers interact in the membrane to allow the transmembrane regions to form a channel through the cytoplasmic membrane. Like pI, pXI spans the membrane. pIV is secreted across the cytoplasmic membrane, and then, integrates into the outer membrane as an oligomer (10 to 12 pIV monomers). pIV oligomers form a large exit pore in the outer membrane through which the assembled phage particle can pass. However, in the absence of phage assembly, it may act as a gate to maintain the integrity of the membrane.

Assembly is a progressive process in which the pV dimers are displaced from, and the appropriate capsid protein added to, the DNA as it is extruded through the bacterial envelope into the media. Assembly continues until the end of the DNA is reached and the phage is released into the media. Throughout this whole process, the membrane must retain its

integrity, as the bacteria continues to grow and divide. The assembly process is divided in 3 parts: initiation, elongation and termination.

Initiation: The substrate for the assembly process is the pV-phage DNA complex. An interaction between the packaging signal and the cytoplasmic domain of pI may initiate assembly. A hypothesis is that pI is able to direct an orderly assembly of pVII, pIX, and pVIII around the packaging signal to form the tip of the particle.

Elongation: The phage is elongated by the processive set of reactions which removes the pV dimers from the phage DNA and replaces them with pVIII as the DNA is extruded through the membrane. These reactions are probably catalyzed by the cytoplasmic portion of pI, and may require ATP hydrolysis. Assembly needs the presence of reduced thioredoxin.

Termination: When the end of the DNA is reached, assembly is terminated by the addition of pVI and pIII.

It is not known whether an assembly site can be used for the synthesis of more than one particle. If only one particle can be assembled in one site, the initiation might be expected to be one of the rate-limiting steps in phage assembly.

The assembly process is tolerated quite well by the host, as the infected bacteria continue to grow and divide with a generation time approximately 50% longer than for an uninfected bacteria. About 1000 phage are produced during the first generation following infection, after which the host produces approximately 100-200 phage per generation. Turbid plaques can be visualized in overlay lawns of *E.coli*, due to the reduced maximal growth of the host cells where phage are produced. Plaques are of varying size and contain around 10^8 infective phage particles (Model and Russel, 1988).

2- Principle of phage display

The membrane-associated assembly of filamentous phage has made them a valuable tool in biological research. The apparent flexibility of the assembly process has led to an impressive array of applications for the use of phage display, where the bacteriophage is used as cloning vehicle. The presence of single-stranded DNA in the particles, allowed a simple and clean substrate for DNA sequencing.

By insertion of coding regions into genes encoding specific phage capsid proteins,

chimeric proteins can be produced which are able to be assembled into phage particles, and are thus displayed on the surface of the filamentous phage particle. Foreign peptides or proteins fused to the pIII capsid proteins have little, if any, influence on the assembly of the capsid proteins around the DNA. Chimeric peptides fused to the pVIII proteins are limited to 6 amino acids, unless particular vectors are used, as described later.

In theory, any protein can be displayed on the surface of the phage, if the following criteria are fulfilled. The foreign protein must be translocated efficiently across the inner membrane into the periplasm so that enough of the protein is available for assembly into particles. Once in the periplasm, it has to be able to fold correctly in the oxidative environment encountered in the periplasm, as well as being able to enter the assembly site and not interfere with the processes that occur at this site during assembly.

George Smith is the first to have displayed a protein on a phage surface. He fused a part of the gene encoding the *Eco*RI endonuclease to the pIII protein (Smith, 1985), and then demonstrated that phage containing the *Eco*RI-pIII fusion protein could be enriched more than 1000 fold from a mixture containing wild type phage, when selected with an immobilized polyclonal antibody specific for the *Eco*RI endonuclease. From this work, it has been concluded that using recombinant DNA technology, it should be possible to build large libraries ($>10^8$) where each phage displays a unique random peptide, and that this methodology provides a direct link between phenotype and genotype.

Two aspects of phage display are described here in more detail: the vectors which are used to display the foreign proteins, or random peptides, and the affinity selection methods applied, to select phage displaying elements which have the desirable features, e.g. affinity to a particular target.

2-1 Vectors for phage display

Many vectors have been designed for displaying proteins or peptides on the surface of phage. They can be classified into two distinct categories: the phage and phagemid vectors.

2-1-1 Phagemid vectors

By combining the features of phage and plasmids, new cloning vectors, called phagemids have been constructed. They contain the replication origin and packaging signal of the filamentous phage together with the plasmid origin of replication, to maintain themselves in the absence of phage proteins, in *E.coli* host, and gene expression systems, to ensure

efficient transcription and translation of the gene-fusion of interest. Except for the capsid phage protein used as fusion partner for the protein or peptide to be displayed, they do not contain any phage genes. They are, thus, unable to replicate as phage-like particles by themselves. After transduction or transformation of bacteria with phagemid DNA, they maintain themselves as plasmids. This feature give a strong stability advantage to the DNA, compared to phage DNA. Indeed, when the DNA replicates as a plasmid, it is double-stranded, while the phage DNA, mainly propagates as single-stranded DNA, where it is observed to be subject to frequent deletions, and where those smaller phage have a selective advantage due to shorter replication time.

To be packaged into phage particles, phagemids require superinfection with a filamentous helper phage (M13KO7), which activates the phage replication origin. The resulting phagemid single-stranded DNA is encapsulated into phage-like particles using phage proteins, provided by the helper phage. The phagemid genome is packaged and secreted more efficiently than the helper phage genome which is defective in replication.

Approximately 90% of the secreted phage particles carry the phagemid genome. All particles display a mixture of wild-type and fusion proteins. The number of fusion proteins displayed, depends on the type of phage protein used for the fusion. This system provides a direct physical link between the protein exposed on the surface, and the genome packaged into the phagemid particles. Since particles from a single clone are assembled in a single cell, a variant from a library maintains coupling between that particular variant protein, and the gene for that variant, although occasional deviations may arise in the case of multiple infections in a single host cell.

2-1-2 Phage vectors

A bacteriophage vector contains a bacteriophage replication origin, along with all phage functions required for viral propagation. That means that it is able to replicate and be assembled as phage particle on its own. Since the phage protein gene used for the fusion is not present as a wild type gene, the phage exposes on its surface only fusion type from this protein. Attempts failed to display some elements on the surface of M13 phage, failed due to defects in viral assembly, stability, and infectivity.

An alternative type a vector uses two copies of one of the capsid protein genes one of which can be used as fusion partner. One copy is the wild type gene, while the second is fused to the protein or peptide to be displayed. Bacterial cells infected with these phage incorporate

both wild-type and fusion copies of the gene products into the same viral particles. These phage vectors have the advantage that they do not require a helper phage for the infectious process (presence of all wild type proteins necessary for the infection, replication and assembly), and display a lower number of fused proteins, in this latter respect, they are comparable to the phagemid system.

2-1-3 Phage genes for the fusion with the protein to be displayed

The two commonly used phage proteins for the fusion, are the pIII and pVIII capsid proteins. Viral vectors that accept and display fusions for genes III and VIII have been termed type 3 and 8. In several vectors, the displayed element is separated from the remainder of pIII or pVIII, by short linkers.

Some proteins have been successfully fused to the carboxyl end of the pVI protein, the vectors generated for the display of pVI-fusion proteins are referred as type 6 (Jespers et al., 1995). But they will not be further discussed here.

The vectors are classified under six nominations, according to the system of fusion protein they deal with (Smith, 1993). Type 3 represents phage vectors which expose the foreign proteins fused to the pIII proteins. All five copies of the pIII display the foreign protein (figure I-3). The type 33 stands for phage vectors which express a hybrid pIII, in addition to a wild type pIII protein. The virion displays a mixture of pIII molecules, only some of which are fused to the foreign protein. And finally, the type 3+3 represents the phagemid system. The vector contains only the pIII gene from the phage, which is fused to the foreign protein. In this system, when the superinfection with the helper phage takes place, wild type (from the helper phage), and recombinant pIII (from the phagemid), are assembled together, leading to a monovalent presentation of the recombinant pIII (Röttgen and Collins, 1995). During the assembly, both phagemid and helper phage are packaged at the same time, and both display the hybrid pIII.

Type 8, 88, and 8+8 vectors are the gene VIII counterparts of the gene III vectors.

The display of peptides or proteins in M13 can have a valency of one (e.g. 3+3 type) to thousands (e.g. type 8) of copies. This is dependant on the display site and the type of vector. The type of fusion, expression levels, and experiment will dictate which display system is appropriate.

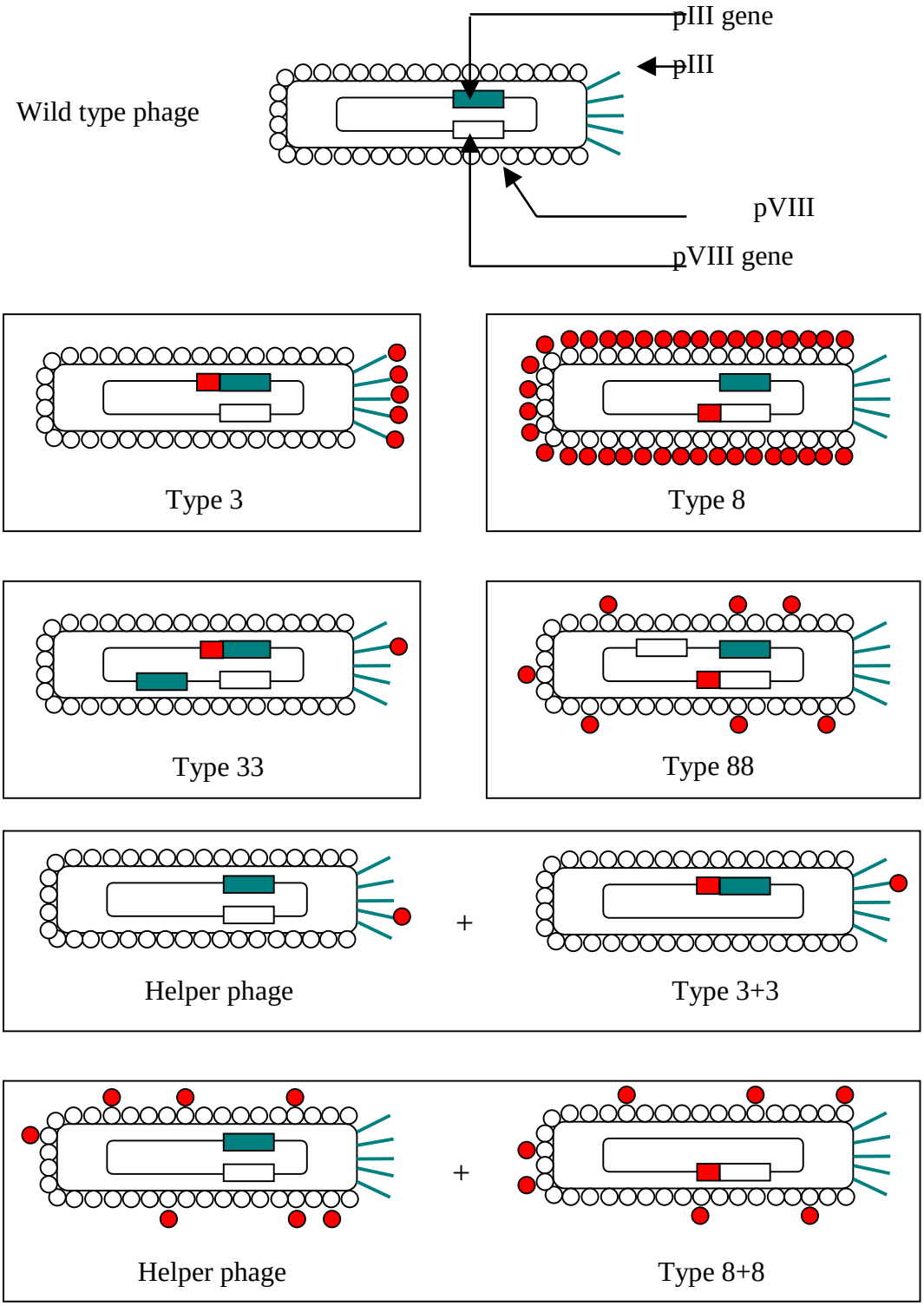


Figure I-3: Comparison of different phage display vector system. The red boxes and spheres correspond to the foreign genetic elements and their encoded proteins, respectively. The vector system is classified as 3 or 8, depending on whether the foreign element is fused to pIII or pVIII proteins. All types possess a single copy of the fusion protein, in addition, types 33 and 88 contain a further copy of the gene used as fusion gene, but this additional copy represents a wild type gene.

Many short peptides and a variety of proteins have been displayed at the N-terminus of mature pIII. Protein domains are more accessible when linked to domain II of pIII. Most antibody display experiments have utilized this truncated form of pIII for efficient display. Replacement of domain I of pIII with an exogenous sequence leads to noninfectious phage. Thus, these particles must also carry a wild-type pIII, expressed from a second gene III in the vector (type 33) or a helper phage genome (type 3+3).

One major feature of the 33 and 3+3 vector systems, is that the displayed element is often monovalent, compared with the type 3 vector system which exhibits 5 fusion particles. In 33 and 3+3 systems, both helper and phagemid/phage viral particles will have 0-5 copies of the pIII fusion protein per particle (this can be advantageous when the pIII fusions yield phage which are not stable of infectious). The reduced valency is also valuable in eliminating avidity, thus permitting discrimination between low- and high- affinity display elements.

Multivalent display	**Monovalent or reduced display**
Type 3 and 8 phage	Type 3+3, 8+8 phagemids, 33, 88 phage
Displayed element might affect phage infectivity.	Negative effect of the displayed element is minimized
Isolation of modest affinity binders **Valency**: 3-5 (pIII) or 2700 (pVIII) fusion proteins (pVIII fusion: limited to short peptides)	Isolation of high affinity binders **Valency**: 0-1 (pIII) or 25-100 (pVIII) fusion proteins (size of the fusion proteins unlimited, also for pVIII)

Table I-1: Comparison of properties of monovalent and polyvalent displays. Monovalent display can be used to discriminate between modest and high affinity binding peptides. Monovalent display is also a tool for the display of variants that may not be tolerated by M13 in 5 to 2700 copies. It has been pointed out by Armstrong (1996), that the selection of weak binders with polyvalent vectors yields sequence (consensus) information which may lead to the development of high affinity ligands.

pVIII appears to tolerate only short inserts of 5 or 6 amino acids additional residues, in the type 8 vectors, but however, can tolerate foreign domains as large as 50 kDa, in type 88, and type 8+8 vectors.

In the 88 and 8+8 vector systems, the valency of the displayed element, is between 25 and 100 copies per viral particle (Kang et al., 1991a). The viral particles have a mixture of both fusion and wild-type pVIII. The increased valency of pVIII display, as compared to pIII, has advantages in experiments in which low-affinity receptors are the target, and has been proposed as an ideal vehicle for developing vaccines where the entire phage particle may be particularly antigenic.

Monovalent or polyvalent displays present different features, and the method of choice has to be defined for each particular target. Their properties are summarized in the Table I-1.

2-1-4 Other viral display vectors

Other vectors have been successfully used to display element on their surface. Among them, bacteriophage λ is commonly used. Proteins can be fused at the C-terminus of its protein V. This is a useful vector for the display of cDNA segments, because large recombinant libraries can be generated in this vector.

Recently, vectors based on the *E.coli* phage T7 have been developed for the display of peptide libraries. Phage T7 has an icosahedral symmetry, and the peptides are displayed at the C terminus of the major coat protein (Houshmand et al., 1999).

Other viruses remain to be exploited, such as RNA phage Qβ, baculovirus, and retroviruses.

2-1-5 Genetic markers

It is essential to use vectors carrying genetic markers, for example, in order to allow a selection of the cells which contain the phage. There are a number of convenient markers, which include antibiotic resistance, epitope expression, *lac*Z expression, and supressor genes.

In a number of vectors, drug resistance genes have been introduced by bacterial transposons. A selectable marker (e.g. drug resistance factor) allows only plasmid-containing cells to survive.

Several vectors have been engineered to express short peptide ”tags” that can be recognized by particular monoclonal antibodies, e.g. the *c-myc* epitope. In some vectors, the

c-myc epitope is in the stuffer fragment where it is useful in immunologically discriminating between parental and recombinant phage particles, for the purposes of estimating the percentage of recombinants in a library.

When a vector carrying *lac*Z gene is in the appropriate bacterial host that has been exposed to the inducer isopropyl-β-D-thiogalactopyranoside (IPTG), functional β-galactosidase activity is present. This gene can easily be detected by the occurrence of blue plaques in bacterial lawns containing the indicator bromo-4-chloro-3-inodocyl-β-D-galactoside (XGal).

In a few vectors, suppressed stop codons have been placed between displayed protein domains and domain II of pIII. When such recombinants are propagated in a suppressor carrying bacterial strain, virus particles are produced incorporating chimeric protein domain pIII into the capsids. After screening phage populations for virus particles with certain properties, individual isolates can be introduced into bacterial strains that lack suppressors. In these hosts, there is efficient translational termination at the stop codon in the chimeric gene III, and consequently, the protein domain is secreted by itself. This is a very efficient way to secrete antibodies (Hoogenboom et al, 1991), or growth hormone (Lowman and Wells, 1993) for testing purposes.

2-2 Screening phage display libraries

The screening of a phage display library is a way to isolate peptide ligands for targets of interest. The bacteriophage expressing binding peptides are selected by affinity purification on a target of interest. The peptide sequences are subsequently deduced from the DNA sequences.

The library screening involves several steps: a suitable library has to be obtained, together with a source of target. Then, the target and the library are incubated together, in order to capture the binding phage. Washing steps remove the non-binding phage. After this, the binding phage are eluted from the immobilized target, and propagated. This cycle has to be repeated several times, until a significant enrichment of the eluted phage is observed. At this stage, individual clones are further analyzed by sequencing, and the clones are compared, with the aim of distinguishing a consensus sequence.

This work focuses on phage display random peptide libraries, as it is the type of bank which is investigated in order to isolate highly specific ligands to our particular targets (SH3 domains).

2-2-1 Important parameters to improve the biopanning

Library quality: the quality of the phage library to screen has an important influence on the probability of isolating phage, binding to a particular target. It is therefore crucial to establish a strategy which will generate an adequate sized bank. The library should, for example, display random peptides, long enough to allow maximization of ligands which cover a large proportion of the surface of the target binding site. However, the display of long random peptides, requires a high complexity of the bank, in order to be diverse enough to contain all possible peptide sequences. The table I-2 gives an idea of the bank diversity necessary to be obtained, according to the random peptide size that they display.

Positions randomized (n)	Peptide diversity (20^n)	DNA diversity (NNK codons) (32^n)	Transformants required for complete library with: 90% confidence	99% confidence
1	20	32	74	149
2	400	$1.0x10^3$	$2.4x10^3$	$4.8x10^3$
3	$8.0x10^3$	$3.3x10^4$	$7.6x10^4$	$1.5x10^5$
4	$1.6x10^5$	$1.1x10^6$	$2.4x10^6$	$4.9x10^6$
5	$3.2x10^6$	$3.4x10^7$	$7.7x10^7$	$1.6x10^8$
6	$6.4x10^7$	$1.1x10^9$	$2.5x10^9$	$5.0x10^9$
7	$1.3x10^9$	$3.4x10^{10}$	$7.9x10^{10}$	$1.6x10^{11}$
8	$2.6x10^{10}$	$1.1x10^{12}$	$2.5\ x10^{12}$	$5.1x10^{12}$

Table I-2: Theoretical required diversity of phage display libraries to ensure the presentation of all possible random peptides. NNK codons (where N is any nucleotide, and K is G or T) is commonly used for randomization in biological libraries. NNK encodes all the amino acids except two of the stop codons, reducing the stop codons occurrence of two third. The confidence that a library contains all possible amino acid sequences is calculated assuming a Poisson distribution (Lowman and Wells, 1991). Adapted from Clackson and Wells observations (1994).

For a phage library displaying 8 amino acid long peptides on the phage surface, a diversity of at least 2.5×10^{12} variants would be necessary to represent every possible sequence (with a confidence of 90%). However, the largest combinatorial peptide libraries which have been generated, contain about 10^{11} variants. Therefore, many laboratories are focusing on developing methods able to overcome this limitation. The generation of recombination within the hypervariable region of the libraries is a powerful concept which is exploited to increase the bank diversity. The DNA shuffling, in use for instance by Stemmer's

group (Stemmer, 1993a/b), and the cosmix-plexing® approach[1], developed in our laboratory (Collins and Röttgen, 1997a, 1997b), demonstrate the advantages of this idea. The advantages and drawbacks of these methods will be discussed later on.

The diversity of the bank might be, in some cases, overestimated. Indeed, in some libraries, the frequency of the number of non-recombinant parental clones, which do not present any peptide, is high. In the worst case, the library would not contain enough variants to be considered useful (e.g. $<10^7$). It is important to be able to test for their presence, or frequency, to have a correct evaluation of the complexity of the bank. In any case, it is crucial to prevent or limit their presence. At the very least, empirical selection experiments should be performed to correlate the results with theoretical expectations. Several methods have been applied to create high quality libraries. Some of them are discussed below:

One technique is to destroy, intentionally, the reading frame of genes III or VIII by inserting a stuffer fragment in the vector. The presence of a stuffer fragment also allows a closer examination of the restriction digests ensuring complete cleavage of the vector with both enzymes at the site at which the hypervariable cassette is to be integrated. Thus, pIII or pVIII proteins are not expressed anymore. The reading frame is restored in the vector when the stuffer fragments are replaced with DNA encoding peptides or protein domains which have the appropriate reading frame. As pIII and pVIII are required components of the virus particle, this strategy permits positive selection of recombinant versus parental genomes (with phage vectors). One caveat of this selection scheme, however, is that reversion of the defective genes occurs frequently, leading to a loss of discrimination between recombinant and parental genomes.

Another means of selecting recombinant versus parental phage, is to include a stop codon in the stuffer fragments of genes III or VIII. The amber stop codon (TAG) can be suppressed efficiently in bacteria containing *sup*E or *sup*F. When these vectors are propagated in such bacterial strains either a glutamine or tyrosine is inserted at the TAG codon, respectively. However, when TAG containing genes III or pVIII are in bacterial strains that lack either *sup*E or *sup*F, no full length pIII or pVIII accumulate, and no virus particles are generated. Almost no parental phage have been found, when ligation mixes are introduced into suppressor-less F' bacteria.

Another alternative which will be described later on, is used in our laboratory. In our hands, absolutely no parental phage have been isolated, using this method.

[1] Cosmix-plexing® is a trademark belonging to Cosmix molecular biologicals GmbH (patent pending)

When checking the bank quality by sequencing random clones to estimate the frequency of non-recombinant phage, parental phage has to be taken into account, along with phage displaying variants containing stop codons, or variants which show changes in the open reading frame, due to point mutations. A high frequency of these three types of defective clones has to be reduced to the minimum, in order to maintain of the complexity of the libraries.

The libraries should remain stable, even when greatly amplified. In general, the phagemid vectors are considered more stable than phage libraries, as the latter depend almost entirely on single-stranded DNA propagation during amplification, which frequently generate small deletion fragments which overgrow the rest of the library. In addition, problems may occur due to intolerance by *E.coli* for certain sequences during DNA replication, e.g. with vectors containing deletion of repeats and palindromes (Collins, 81; Collins et al, 82). The intolerance of particular gene products can be reduced during propagation, by using strongly repressible promoters, such as λp_L, for the control of production of the protein or peptide to be presented on the phage (Courtney et al, 95; Röttgen and Collins, 95; Maenaka et al, 96). The displayed elements might also be wrongly folded.

To minimize problems arising from instability of phage displayed peptides, phage should be protected from proteolytic degradation. Therefore, phage used in screening experiments should be from a **freshly amplified stock**.

The target quality is also an essential parameter influencing the success of the panning. The target must retain its native conformation during affinity purification of binding phage, at least during the initial binding and the washing steps. When possible, it is of interest to have a functional activity test for each particular target. In addition, for standard libraries, the target has to be highly purified (this not always necessary for cosmix-plexing® libraries, where successful selections have also been carried out on complex targets). Otherwise, the clones that bind to the target may be discriminated from those that interact with contaminants in binding reaction. Furthermore, the panning procedure seems to depend to a great extent on the way in which the target is presented (e.g. immobilized) to the phage library, during the affinity selection.

The most common strategy to immobilize a target is to adsorb it onto ELISA-treated microtiter plates. Passive adsorption is typically sufficient to immobilize the modest quantities of target required for successful isolation of binding phage. Using this format, a single well is

generally required per round of purification per target/library combination. However, the target may be immobilized onto many other solid supports such as Sepharose (Dedman et al, 93), or paramagnetic beads. Immobilization on paramagnetic beads offers the potential advantages of increased surface area, and therefore target concentration per unit volume, relative to target immobilization onto microtiter plates, although a larger amount of protein is typically required for immobilization (it should be noted that target concentration is the driving force for the binding kinetics, at least in the initial selection round. Target concentration becomes less important in later rounds and can be deliberately reduced as a factor in increasing stringency). A strong magnetic source is used to draw the target-bead-binding phage complexes during washing steps.

Some proteins may become denaturated and lose binding activity when bound to plastic or solid supports. These problems may be circumvented by using an intermediate capture protein to immobilize the target, such as immobilized protein A can be used to capture an antibody via its Fc while still allowing access to its Fab, or immobilized anti-target antibodies can be used to capture the target. It is critical, when using this capture method, to minimize recovery of phage which bind to the capture protein rather than to the target protein.

2-2-2 Affinity selection

Initial challenge: the affinity selection allows isolation of variants from a phage library, which exhibit required properties, e.g. high binding affinity to a particular target molecule. This strategy is often referred as “panning”, in analogy with the procedure of panning for gold dust by shaking river sediment in flowing water. The initial number of phage used from a library for a first round of panning selection is typically about 10^{11}-10^{12} cfu.

The affinity selection of the phage library can be performed using two different approaches. In one case, the phage library is incubated with preimmobilized target, such that the binding phage are temporarily immobilized onto solid-phase while non-binding phage are washed away.

In the alternative approach, the phage library is incubated with target in solution, and subsequently captured as phage-target complexes onto solid phase while the non-binding phage can be washed away.

Washing steps: after the incubation between the target and the phage library, intensive washing steps are achieved in order to wash away the particles showing low affinity for the target, leaving the strongly binding particles attached to the target. This allows

discrimination between weak and strong binding particles.

Elution: the enriched phage population, remaining bound to the target after the stringent washing is then eluted from the target-phage complexes. This is an essential phase of the panning procedure. The most commonly used method, is to elute the phage with drastic pH changes (i.e., glycine-HCl, pH 2.2), which result in target or peptide denaturation without loss of phage infectivity.

An alternative procedure is the elution by treatment with 6M guanidine-HCl, 2M urea, or 10 mM DTT, although these reagents must be removed prior to phage amplification. The use of 1% SDS or phenol to elute the binding particles is also possible, although, it requires DNA transformation to regenerate infectious phage particles.

Another means is to take advantage of particular binding characteristics of the target. For example, if the target needs a metal ion for binding, phage elution might be successful with EDTA, or EGTA.

Another kind of elution is the use of natural or synthetic ligands for competitive elution from receptors, or interacting proteins for competitive displacement by formation of multimeric complexes. This has the advantage of selecting for specificity for the site at which the selected ligands binds. Specificity can also be increased by adding competitive targets (soluble) during the initial affinity selection steps.

A simple and reliable elution method which can always be used in the absence of knowledge of an appropriate elution method, is to add the *E.coli* host cells directly to the immobilized complexes after the washing steps.

Amplification: the eluted phage, still biologically viable, can reinfect a new *E.coli* host in which it propagates, leading to amplification. Despite the high number of phage used in the first round of selection (10^{11}-10^{12} cfu), the number of phage particles representing any given binding clone that are recovered from the first round is reduced to the point that binding clones may be lost if subjected to a second round of purification without intervening amplification.

The amplification should result in a 10^5 to 10^7 fold increase in the titer of any given clone from the previous round of affinity selection. Every single clone enriched in the first selection round, participates in the further rounds of selection at a greatly increased concentration.

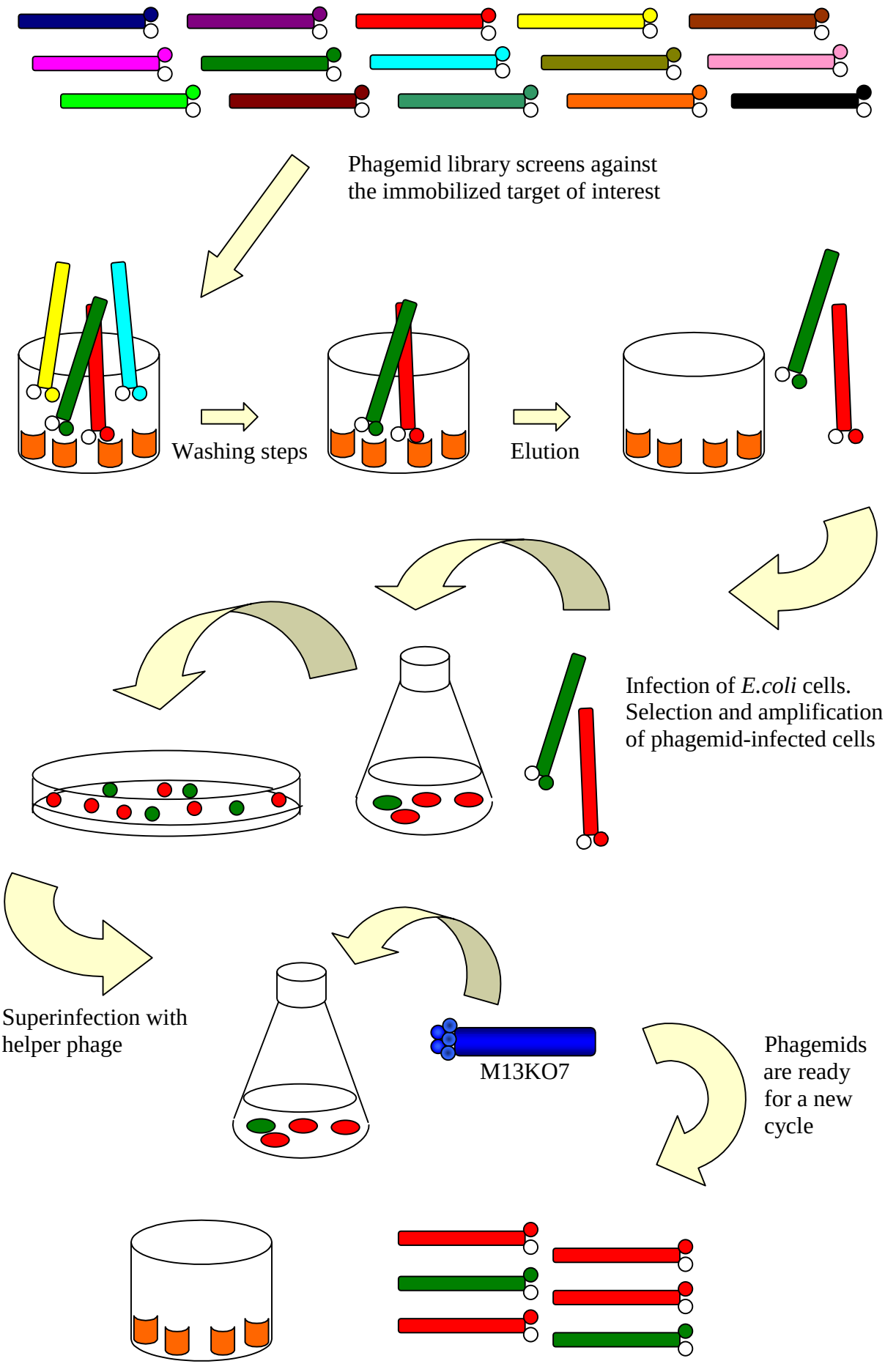

Phagemid library screens against
the immobilized target of interest
Washing steps
Elution
Infection of *E.coli* cells.
Selection and amplification
of phagemid-infected cells
Superinfection with
helper phage
M13KO7
Phagemids
are ready
for a new
cycle

Figure I-4: The affinity selection cycle. Phage particles presenting the various protein or peptide variants are incubated with the immobilized target molecule. Particles with low affinity for the target are washed away. The remaining particles are eluted and fresh *E.coli* cells are infected with them. Clones are selected for antibiotic resistance. Particles are collected, and packaged, by superinfection with helper phage. The enriched clones are now ready to be screened in the next cycle.

Phage amplification may be performed on petri plates or in liquid culture. It is suggested that there is little difference in the growth of phage by either method (McConnell et al, 95). Liquid culture can accommodate a higher amount of input phage than amplification on plates.

One round of affinity selection with a target immobilized on ELISA plates, typically results in an 10^3 to 10^5 fold enrichment of binding over non-binding phage. Thus, three or four rounds of panning are generally sufficient for the isolation of a small number of clones containing the phage with the highest binding affinities.

The panning protocol described in figure I-4 is conceived for a phagemid random peptide library, taking advantages of the methods using preimmobilized target, acidic elution of the binding phage, and plate amplification of the library between each round of affinity purification.

3- Applications

The ability to construct libraries of enormous molecular diversity, having a direct link between phenotype and genotype, in combination with selection for molecules with predetermined properties, has made this technology applicable to a wide range of problems.

3-1 Phage display of natural peptides

The main application of displayed natural peptides, is for the epitope mapping of monoclonal, or polyclonal antibodies. In this case, phage display is seen as a means of identifying clones that were expressing epitopes recognized by a given antibody.

Traditionally, epitope mapping of protein antigens has relied on heavy physical chemical analysis: First, the purified antigen was fragmented with proteases, followed by identification and sequencing of the reactive fragments. Then, chemical modification experiments were performed, in which residues interacting with the antibody are protected from modification. A series of peptides corresponding to the primary structure of the antigen have to be synthesized. And finally, a direct physical characterization using NMR or X-ray

crystallography has to be achieved.

As an alternative to this method, phage display can be used to localize the antigenic epitope relatively quickly. Fragments of DNA that encode portions of the protein antigen are fused to the gene encoding capsid protein III. And phage can then be tested with the antibody to determine which displayed fragments react with it (Tsunetsugu-Yokota et al., 1991; Wang et al, 1995).

The limitation of this strategy is that each antigen represents the construction of a new phage library. Furthermore, the starting point, DNA encoding the antigen must encode the epitope.

The display of natural peptides can also be used with the aim of generating immunogens. Short segments of various proteins have been displayed on M13 virus particles for the purposes of eliciting antibodies against the coat proteins of certain parasites and viruses (De La Cruz et al., 1988; Greenwood et al., 1991).

3-2 Phage display of random peptides

Combinatorial peptide libraries are an important application of phage display. Synthetic oligonucleotides, fixed in length but with unspecified codons, can be cloned as fusions to genes III or VIII of M13 where they are expressed as capsid fusion proteins.

These random peptide libraries can be tested for binding to target molecules of interest, due to affinity selection. The library is incubated with the immobilized target molecule, followed by washing procedure to remove the weak or non binders. Finally, the bound phage are eluted, and amplified, before being again selected for binding to the target molecule, thus enriching those phage that bind strongly to it. Thus rare phage that bind can easily be selected from greater than 10^8 different individuals in one experiment.

The displayed peptides can have a free amino terminus, thus allowing the peptide considerable flexibility much like peptides free in solution.

Random peptide libraries have been invaluable in mapping the specificity of the antibody binding sites. They represent a source of sequences from which epitopes and mimotopes can be operationally defined (Geysen et al., 1986).

Random peptide libraries are a very powerful tool to identify peptide ligands. This way of working with phage display is the one which was chosen for this work. Random

peptide libraries are a rich source of peptide ligands for a variety of receptors. They also provide antagonists for protein-protein (as instance, on Src SH3 domain: Cheadle et al., 1994; Rickles et al., 1994; Sparks et al., 1994) and protein-nonpeptide ligand interaction (Krook et al. (1994) identified peptide ligands against single-stranded DNA), *in vitro* and *in vivo*.

Another important use of random peptide libraries has been the development of "substrate phage" (Matthews et al., 1993). In this case, the library is used as a means of defining substrate specificity rather than simply binding to a target molecule. In theory, many different post-translational modifications (i.e., phosphorylation, glycosylation...) can be mapped the same way.

3-3 Phage display of protein and protein domains

In many cases, the phage displayed protein retains its normal binding or enzymatic activity even when fused to the N-terminus of mature pIII or pVIII phage proteins.

The coding region for a protein to be displayed on surface of a phage, can be mutagenized by cassette mutagenesis, error-prone PCR, or shuffling to generate a plurality of altered sequences. This is a way to direct evolution, and from the mutagenized population, it is possible to select stronger binding ligands to a particular target (as instance, Lowman et al. 1991; Dennis and Lazarus, 1994).

One of the more powerful applications of phage display has been in the area of antibody engineering. It is possible to express both Fab and single-chain Fv (scFv) antibody fragments on the surface of M13 viral particle with no apparent loss of the antibody's affinity and specificity. The coding region of V_L and V_H chains can be obtained from naive mice (Gram et al., 1991), immunized mice (Clackson et al., 1991), or germline genes (Hoogenboom and Winter, 1992). High affinity antibodies can be isolated in this way.

It is also possible to express cDNA-encoded proteins on the surface of phage which can then be tested against a particular immobilized target *in vitro*, using biopanning enrichment (Crameri and Suter, 1993). This allows the study of signal transduction pathways, in which a whole series of protein domains interact with one another.

Direct selection of protein-protein interactions has also proven possible with phage-displayed libraries. For M13 phage particles to be infective both amino and carboxy domains

of pIII must be present. It is possible to express the target protein as a fusion protein with the carboxy-terminal half of pIII separately from cDNA segments which have been expressed and fused to the amino terminal domain of pIII. When the target and particular cDNA-expressed protein domains interact, the two pIII domains are brought together and the resulting phage particles regain infectivity (Gramatikoff et al., 1994; Duenas and Borrebaeck, 1994). This system complements the yeast two hybrid system.

C – Optimization of binding capacities of interacting partners

A close look at the results achieved in studies of protein-protein interactions shows how crucial it is, to characterize interacting partners with optimized affinity and/or specificity. Several kind of approaches have been developed to isolate more strongly interacting partners. These can be classified into two groups. The first tends to ameliorate binding capacity of already existing ligands (so called, natural ligands). The second strategy, on the contrary, concentrates on isolating novel ligands (through phage display for example), and focuses on developing alternatives to increase their binding capacities to the target molecule.

1- Natural ligands

Natural ligands are found in nature, and are able to make interactions with other partners also existing in nature.

1-1 Consolidated ligands

In this approach, Cowburn et al. (1995), created a series of bivalent ligands, which interact simultaneously with two different domains of the same protein. They chose as model the SH2/SH3 domains, since it is very common to find both of them within one protein. The Abl protein (Abelson protein-tyrosine kinase) was taken for these experiments, as it contains, in addition to a kinase domain, a SH2 domain and a SH3 domain. Furthermore, natural ligands are well characterized for each of these domains. The method is based on the synthesis of “bivalent consolidated ligands”, which are formed through the association between both natural ligands of SH2 and SH3 domains of Abl. The two ligands are separated by a linker, and the bivalent ligand is expected to bind with higher affinity and specificity to Abl, than the isolated ligands, when the linker between the two affine segments is of optimal length. This approach takes advantage of a potential cooperative effect of both ligands. Cowburn successfully shows that such bivalent consolidated ligands present an increased

affinity of approximately two orders of magnitude (Cowburn et al., 1995).

1-2 Rational design

The rational design strategy, proposed by Pisabarro and Serrano, is based on computer predictions (due to GRID and LUDI computer programs), to select for mutations which might enhance the binding capacity of two interacting partners (Pisabarro and Serrano, 1996).

They also took as model the Abl protein, and attempt to improve the affinity and specificity of Abl-SH3 domain natural ligands. Therefore, they selected a natural ligand which shows cross reaction with the Fyn SH3 domain, and attempted to design a ligand which discriminates between Abl and Fyn SH3 domains. Two strategies have been used.

The first one takes advantage of the fact that SH3 ligands have to form a polyproline type II helix to be able to bind to SH3 domains. Thus, the tendency of the natural ligand to adopt such a polyproline type II helix, was increased, by mutation of non proline residues, one after the other, to proline, so that the entropic cost upon formation of the helix would be smaller. This strategy might increase the affinity, but not the specificity.

The second approach is using the GRID and LUDI computer programs to identify reasonable substitutions which might create favorable interactions, more specific to Abl-SH3 domain. Substitutions might occur which fill up the space between the two partners, in a similar way, but which might generate stabilizing interactions such as hydrogen bonds between the two interacting partners.

By using these strategies, Pisabarro and Serrano succeeded to design a peptide which differs from the original ligand at only two positions (P_4-Y, and L_8-P), but which binds the Abl-SH3 100 fold higher than the original natural ligands (K_D of 0.4 µM was reached), and has 1000 fold more affinity for Abl, than for Fyn SH3 domain.

2-Novel ligands – increasing the diversity of the banks

One factor limiting the success of phage display technology is the size of the library. To maximize interactions between the selected ligand and the target to which it binds, it would be desirable to display relatively long randomized peptides. However, the minimal number of clones required within a library (in order to contain all possible sequences) increases exponentially with the number of amino acid residues in the variant peptide. A phage library which displays a 9 amino acid long peptide on its surface, would require a population of at least $1.5x10^{14}$ variants, to ensure a confidence level of 99% that all possible variants are presented ($1.6x10^{20}$ for a 13 amino acid long peptide). However, the largest

combinatorial peptide libraries which have been constructed contain about 10^{11} variants, thus only a fraction of all possible variants are represented in the library. Therefore, methods to minimize this problem have been developed.

2-1 Generating more primary clones

The first parameter to be considered in increasing the library diversity, is the number of clones obtained from the electroporation step. Indeed, the dependency on the electroporation steps, seriously limits the efficiency of primary bank formation (10^5 to 10^6 clones per μg of ligated plasmid DNA, when using a standard cloning method to insert the variants within the vector). Therefore, λ carrier vectors have been constructed which allow the efficient use of the λ *in vitro* packaging system for introducing the phagemid display vector into the cell via transduction. Since the phagemid vector is bracketed by the packaging-start and-stop sequences of bacteriophage M13, superinfection of the transduced λ-phagemid hybrid with an M13-type helper phage causes the phagemid genome to be efficiently excised (Short et al., 1988), packaged, and secreted in the culture medium. The efficiency of the λ *in vitro* packaging is some 4 fold higher than the efficiencies quoted for electroporation, namely 6 to 7.10^7 pfu/μg of DNA (Hogrefe et al., 1993).

A novel protocol is optimized in which a cosmid vector is used to package and efficiently transduce up to 9 phagemid genomes at a time. The vectors used in this system carry a λ cos site, which is recognized by the λ phage, and which allows the packaging within the lambda prehead. These are then resolved to individual phagemid genomes upon superinfection of the recipient bacterium with M13 helper phage (Collins et al., unpublished data).

Another route to increase the number of primary clones obtained by electroporation, is to optimize the method used for the cloning of the variant inserts within the vectors. Such an approach was applied in our laboratory, reaching a very high electroporation efficiency, over 10^8 colonies/μg of ligated DNA (described on the results chapter).

2-2 Create new clones from the primary library

By using different approaches, it is possible to modify the variants of a primary library to create a more diverse or modified secondary library. Several groups have been involved in the attempt to elaborate strategies which try to reproduce the process of natural evolution

(Gallop et al., 1994). The elements involved are mutation and recombination, and in many cases, are simultaneously applied.

2-2-1 Introducing mutations

One way to increase the size of the library is to create mutations within the gene which codes for the peptide to be displayed on the surface of the phage. Several approaches have been successfully attempted to reach this aim.

Error prone PCR

Error prone PCR is commonly used to introduce mutations in the DNA during amplification. It has been used to introduce random mutations into the scaffold of phage-displayed proteins, in an attempt to imitate the introduction of mutations during the somatic maturation of antibodies. DNA synthesis *in vitro* tends to produce random sequence errors. *Vent* and *Pfu* are less error prone than *Taq* DNA polymerase (Cline et al., 1996). The mutation frequency can be easily increased by modifying the buffer composition (raising the Mg^{2+} concentration, including Mn^{2+}, or increasing the pH) or by other factors, such as, lowering the initial concentration of the template for PCR, increasing the amount of polymerase, using a great number of cycles, a long extension time, or a low annealing temperature. In addition, a biased pool of the four dNTPs, with concentrations differing by a factor of 10 to 1000, also encourages a higher error rate (Leung et al., 1989; Pannekoek et al., 1993; Cline et al., 1996; Ling and Robinson, 1997).

As an example, Krykbaev et al. (1997), built a library of mutant CD4 proteins from which variants with improved binding to the HIV gp120 have been isolated after biopanning. Interestingly all of these selected variants had one or two amino acid substitutions within the V1 domain of CD4, at positions located in the strands surrounding the main binding loop.

Mutator strains

Another method, introduced by Low et al. (1996), is to use modified bacterial strains which bear a proof-reading-negative DNA polymerase III (*E.coli* mutD5 mutator strain). In this strain, the fidelity during replication of the DNA is highly reduced. The mutation rate in these strains is only 50 fold above background in minimal medium, but can be increased by 20 to 2000 fold by growth in rich medium. This approach is applied in our laboratory, where the DNA is propagated into WK6λmutS *E.coli* strain, which is mutated to affect the fidelity in DNA replication. The accumulation of created mutations might generate more diversity, and

might contribute to an increase of the binding affinity of a few variants (not demonstrated).

Mutational libraries: the 70:10:10:10 approach

The mutational library approach developed by Yanofsky et al. (1996), takes advantage of previously isolated ligands from a peptide library (through panning of affinity selection). The idea is that, when a ligand has been identified against a particular target, it is still possible, after modification by mutations, to select, among the new batch of clones, for better binding variants. The principle of this strategy is not to increase the diversity of the library by itself, but instead, to enhance the binding capacities of a previously determined ligand.

The sequence of the original binding peptide is mutated at a rate of 50% per amino acid residue (for amino acids encoded by codons where the third position is completely degenerated; 66% mutation rate for amino acids encoded by a single codon sequence). Three additional completely random amino acids are added, as extensions, at both ends.

This frequency of mutation is achieved by synthesizing the oligonucleotides in which the original base at a particular position, is incorporated at a frequency of 70% and the 3 other bases at 10% each. Therefore, this method is referred to as the 70:10:10:10 approach (Yanofsky et al., 1996). This strategy can be used in optimizing the affinity of a particular ligand. The initially isolated binding ligand is submitted to this 70:10:10:10 approach, further rounds of panning are achieved, from which better binding clones might be characterized.

Alanine scan

The alanine scanning mutagenesis is designed to determine which amino acid residues, within a particular ligand, are critical for high affinity and specificity binding to the target. To reach this aim, individual positions of a peptide ligand are successively substituted by alanine, and the modified peptides are tested for their ability to bind to the target. Their binding capacities are compared to the one of the original ligand. Replacement of certain residues leads to a loss or diminishment of binding, whereas replacement of other positions does not. It is therefore possible to determine the relative importance of each residues within the peptide. This technique is routinely used to confirm the consensus sequence conceived from ligands isolated with phage display technology. Commonly, the positions determined within the consensus sequence are the one where an exchange with an alanine residue leads to the loss of the binding abilities for the target, as shown in the work of Kurakin et al. (1998), or Feng et al. (1995).

However, all of these approaches: error prone PCR, modified bacterial strains, mutational library, or alanine scan methods, have a common limitation. They alter DNA only with respect to the starting DNA.

2-2-2 Recombination

Another way to increase the diversity of a library, is to allow recombination within the gene coding for the displayed variant, or directly within the exhibited protein. The use of recombination techniques to generate reassortment between regions of DNA or proteins represents a very important recent development with respect to phage display technology.

DNA shuffling

A method to create *in vitro* DNA recombination has been proposed and developed by Stemmer, and demonstrates that the generation of diversity via recombination is a very powerful concept. Libraries can be cross combined by using DNA shuffling, also called sexual PCR (Stemmer, 1994a,b).

This method takes advantage of both mutagenesis and recombination approaches, since mutated DNA fragments are reassorted together, in order to obtain novel variants. The DNA shuffling involves a partial digestion of a large gene with DNaseI, to form a pool of random DNA fragments (Figure I-5). These fragments are reassembled into a full length gene, by repeated cycles of annealing in the presence of DNA polymerase. The different overlapping DNA fragments act as primers. The fragments prime each other based on homology, and recombination occurs when fragments from one copy of a gene, prime on another copy, resulting in a template switch. Sequences of at least 10 kb, such as genes, gene libraries, or plasmids, can be reconstituted by DNA shuffling, starting with 100-300 bp fragments obtained by nuclease digestion of the sequence of interest. However, this method cannot be used in all cases. It can work only with families of genes or gene variants which contain a high homology among themselves.

Chain shuffling

Winter´s group proposed the chain shuffling method, to mimic the affinity maturation of antibodies that occurs *in vivo* during the course of an immune response (Clackson et al., 1991; Kang et al., 1991b; Hawkins et al., 1992). *In vitro* affinity improvement is accomplished by selecting a pool of antibodies on an immobilized target which is submitted to variations, and to reapply selection. The variation is produced by shuffling heavy and light

chains to create new chain pairings and by mutagenesing the complementarity determining regions (CDRs) to create new variations of the site (CDR1,-2,-3). Many of the new pairings are likely to give a low affinity, however, a few of them may have an improved affinity.

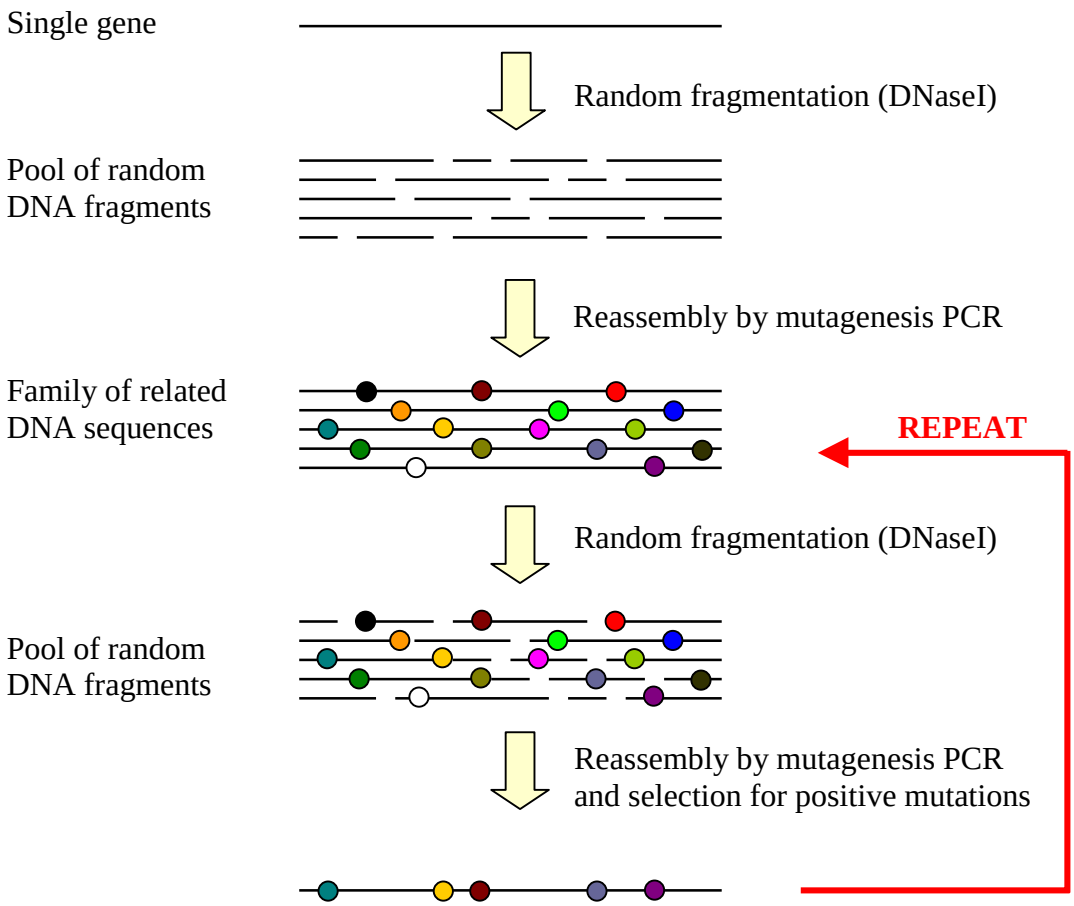


Figure I-5: DNA shuffling: schematic illustration. An original gene is digested with DNaseI, and reassembled through mutagenesis PCR (error prone PCR, as instance). The pool of homologous genes with different point mutations which results from this PCR, is submitted once more to a fragmentation with DNaseI. Reassembling of the random fragments into full-length genes, results in frequent template switching and recombination. Selection from the library of recombinants points out recombinant genes, based on their improved function. The selected pool of improved recombinants provides the starting point for another round of mutation and recombination.

This method has been successfully applied to isolate antibodies with higher affinities against a panel of different targets, among them: the hapten nitrophenyl phophonamidate (Kang et al., 1991b), the hapten phOx (Marks et al., 1992), or the cytomegalovirus glycoprotein B (Ohlin et al., 1996).

Two chain shuffling recombination systems have been described: the *att* and the

Cre/*loxP* systems. They find an application in the generation of hypervariable banks of antibody libraries by chain shuffling of the light and heavy chain subunits.

The att system: this is a λ recombination method, to generate combinatorial libraries, which utilizes the lysogenisation system of λ. λ phage uses a site-specific recombination to integrate into the host chromosome during lysogenisation. The only λ phage function that it requires is the λint gene. The recombination is irreversible in the absence of λxis gene product.

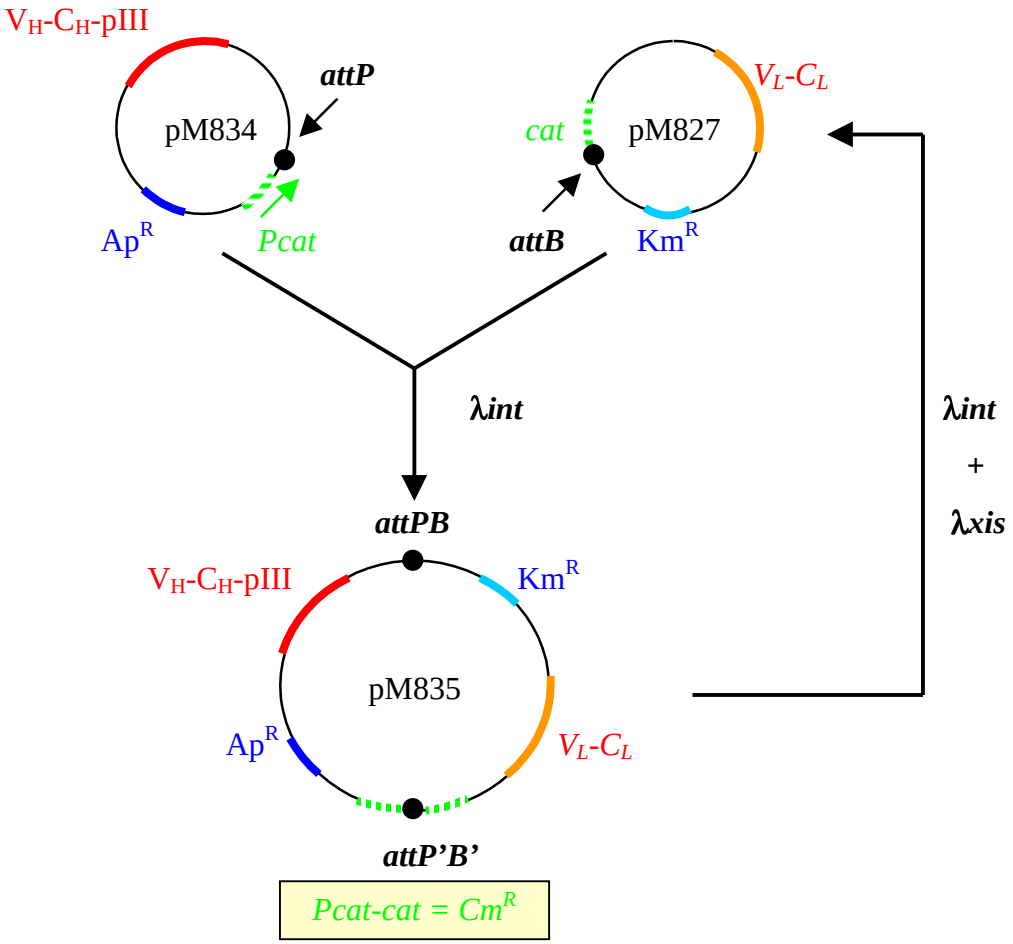


Figure I-6: The *att* system. The phagemid pM834 heavy-chain-pIII fusion library is introduced into the same strain as the light chain library, carried in a low copy number plasmid pM827. *Pcat* refers to the promoter for the chloramphenicol resistance gene. *cat* is the structural gene for chloramphenicol transacetylase without promoter. After the Int-mediated site-specific recombination the required phagemid form pM835 is achieved. Ap^R, Km^R and Cm^R refer to the ampicillin, kanamycin and chloramphenicol resistance genes, respectively. The two sites *attB* and *attP* give site-specific recombination mediated by the phage protein Int, to yield a cointegrate bordered by recombined *att* sites designed *attPB* and *attP'B'*. The reaction can only be reversed if Int and another phage protein Xis are present.

Geoffroy et al. (1994) have described a pair of vectors: a plasmid (pM827) containing *attB*; and a phagemid (pM834) containing *attP* (Figure I-6). On induction of Int protein (heat shock induction of *E.coli* strain D1210HP) the two plasmids recombine irreversibly at these sites, forming a cointegrate, pM835.

pM834 contains the promoter for the chloramphenicol resistance gene (*cat*) near the *att* site, while pM827 carries the promoterless *cat* gene. After recombination the promoter is juxtaposed upstream of the resistance gene, so that the new cointegrate is a phagemid encoding chloramphenicol resistance (only the recombined vector carries a functional chloramphenicol resistance gene). One vector encodes immunoglobulin V_L-C_L and the other encodes V_H-C_H-pIII fusions. The final cointegrate is a phagemid (3+3), specifying chloramphenicol resistance, which will present F_{ab} on packaging, after superinfection with a helper phage. Chain shuffling, and the generation of large combinatorial libraries, can thus be achieved by infecting a heavy chain library on the phagemid pM834 into a population carrying a light chain library in pM827. Each time the bank is created new combinations of heavy and light chains are produced.

The Cre/loxP system derived from the bacteriophage P1, and requires a 33 bp target sequence (*loxP*) which brackets the insert after recombination. Waterhouse et al. (1993), and Griffiths et al. (1994), first demonstrated the use of the Cre/*loxP* mediated gene cassette exchange to shuffle the light and heavy chains in a type 3 library. The phage used in the library encode genes whose products associate to form a Fab fragment presented on the phage, since the heavy chain is fused to the pIII gene. The heavy chain genes are bracketed by a copy of the $loxP_{511}$ and the $loxP_{wt}$ (wild type) sites, allowing exchange of this cassette between two replicons, via Cre mediated site-specific recombination. The mutant only recombines with another mutant *loxP* site, but not with a wild type site, and vice versa. Recombination is induced by infecting the cells carrying the phage, with a P1-*cre* phage, which supplies the Cre recombinase. The efficiency of recombination was estimated to have caused 28% of the acceptor phage to have received a heavy chain from the donor vector (Waterhouse et al., 1993). This acts in a reversible fashion to allow recombination between all homologous *loxP* sites.

The method has also been successfully applied to the chain shuffling of scFv presenting type 3+3 phagemids (Tsurushita et al., 1996). In a recent development, Bradbury (unpublished) used single vector recombination system in which infection of single bacteria by multiple phagemids followed by Cre mediated recombination resulted in the creation of a

library estimated to be greater to 10^{10} in size. This method has the potential advantage that recombination between cycles of selection may result in considerable improvements in affinity, although this was not shown yet.

Libraries generated by Cre/*loxP* recombination are considerably larger than commonly produced (Griffiths et al., 1994, Bradbury, unpublished). In addition, they have been proved to allow isolation of high affinity antibodies to a number of targets.

However, libraries constructed by Cre/*loxP* recombination, where the *loxP* sites are present as direct repeats, might be expected to show instability, due to deletions of the repeats.

The exon shuffling is an ingenious method developed to avoid the problem of having the loxP site (11 amino acids defined) in the reading frame of a gene fusion, when using the Cre/*loxP in vivo* recombination system to shuffle contiguous regions. Fisch et al. (1996) use a self-splicing intron, which removes itself from the mRNA. A *loxP* site has been inserted into the intron, while still preserving the splicing function of the intron. More than 10^{11} clones have been produced using this technique, and highly specific binding clones were obtained on several distinct targets.

A technology developed by Collins and Röttgen is also based on DNA recombination, but overcomes the homology problem faced with the DNA shuffling method. This DNA recombination strategy, leads to the formation of new recombinant variants displayed on the surface of the phage particle, which is the equivalent, for random peptide libraries, of the chain shuffling system, convenient for antibody libraries.

This strategy can lead to an increase of the diversity of the primary library at any time during or after creation of the primary library (by recombination between variants, and/or by introduction of a project specific cassette within the hypervariable domain of a random peptide library). It is also possible to apply this method to a preselected population (i.e., clones recovered after a first round of panning selection), in order to recombine the variants among themselves. During further panning cycles with this recombined library, clones with optimized affinity for the target of interest, can be characterized.

This powerful strategy is the cosmix-plexing® approach.

D – Cosmix-plexing®

The development of a novel technology, termed cosmix-plexing®, allows recombination within hypervariable DNA regions which bear only a two base pair homology. This novel approach has been applied to reassort DNA which encodes random peptides, or antibodies displayed on surface of the phage (Collins, Röttgen patent).

With this method, the potential diversity of the library is no longer represented solely by the number of variants present in the initial library. Libraries made for use with this technique contain a recombination site within the hypervariable domain. Typically, recombination will be carried out with preselected population of (say) 10^4 variants, thus generating 10^8 possible recombinants, containing essentially all the possible combinations of sequences on either side of the recombination site (Collins and Röttgen, 1997b). The advantage of this technique due to the increase in diversity of a preselected population thus achieved, is shown in Figure I-7.

Using the cosmix-plexing® approach, one reassorts the left and right sections of the variable domain, by the use of Type IIs restriction enzymes. The characteristic of these enzymes, is that they bind at a defined sequence, but cleave at a site a fixed distance away, i.e. in a region with no demands on sequence specificity. Thus, a population with "optimized" sequences for a larger number of residues can be generated.

The generation of recombination within a particular region is a powerful concept, as discussed previously with respect to the DNA shuffling system developed by Stemmer (Stemmer, 1994a/b). However, by comparison to the DNA shuffling, our method is not limited by homologies between recombination partners, and shows a higher efficiency for recombination within short regions.

Although the main advantage of cosmix-plexing® induced recombination is obtain through the increase in diversity during selection, it can also be used for the simple production of large libraries (e.g. $>10^9$ variants).

In addition, the production of extension libraries, by the introduction of project specific cassettes into existing libraries, is facilitated (Figure I-8). These cassettes are of particular interest if a "core motif" is known (for example, Pro Pro Xxx Pro motif for the SH3 or EVH1 domains, which show a preference for proline rich ligands).

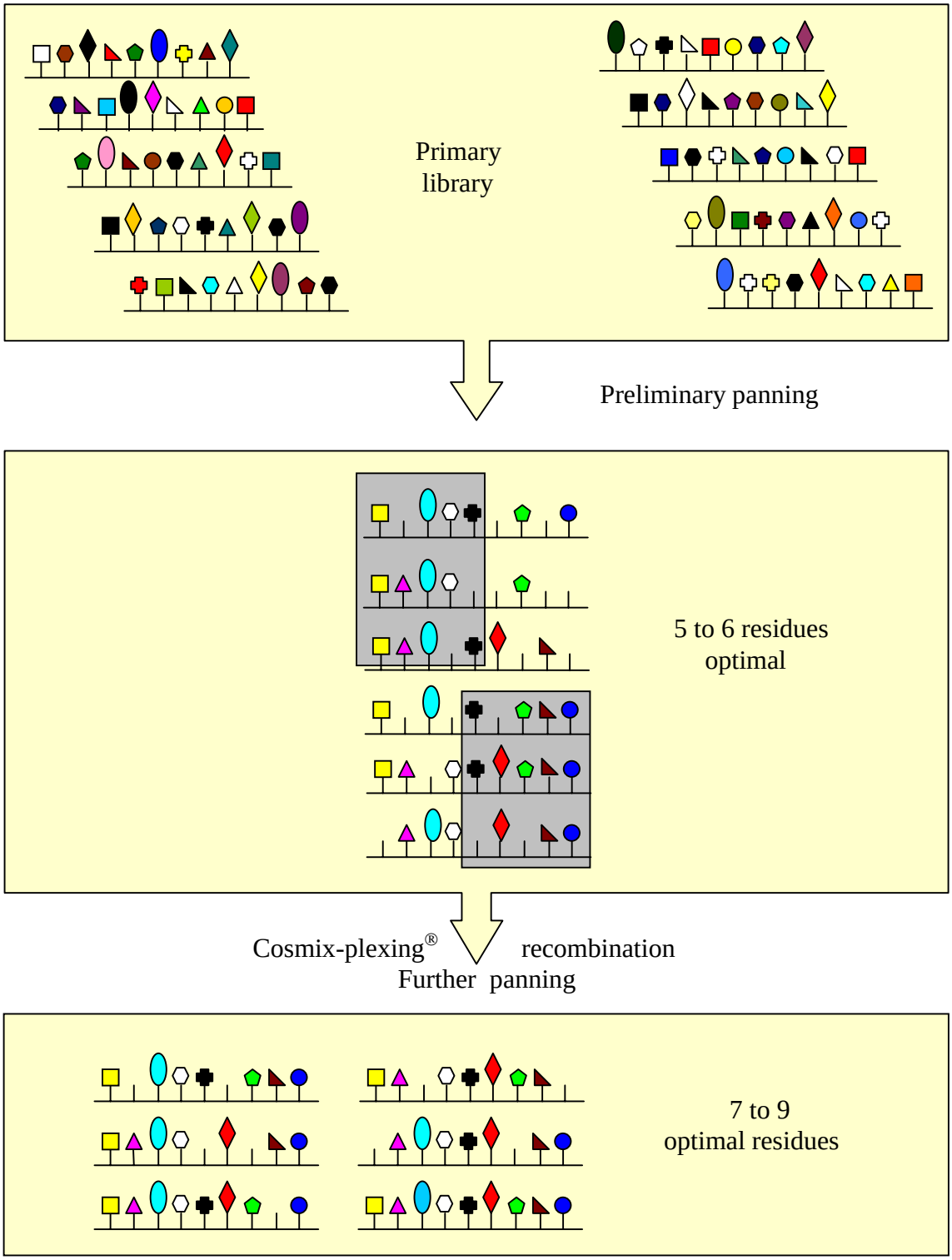


Figure I-7: Cartoon of the principle of cosmix-plexing® induced recombination during screening. A medium-sized combinatorial library (upper section) is subjected to an affinity selection procedure, leading to the isolation of several thousand variants at least partially fulfilling the selected requirement, in which on average some six amino acid residues may be "optimal" (middle section). The left and right section of the hypervariable domain is recombined so that all possible recombinants are produced. Further selection now leads to isolation of "optimal" variants containing eight or nine optimized residues (Theoretically, one can optimized until 12 residues, if the sequence is that long; lower section).

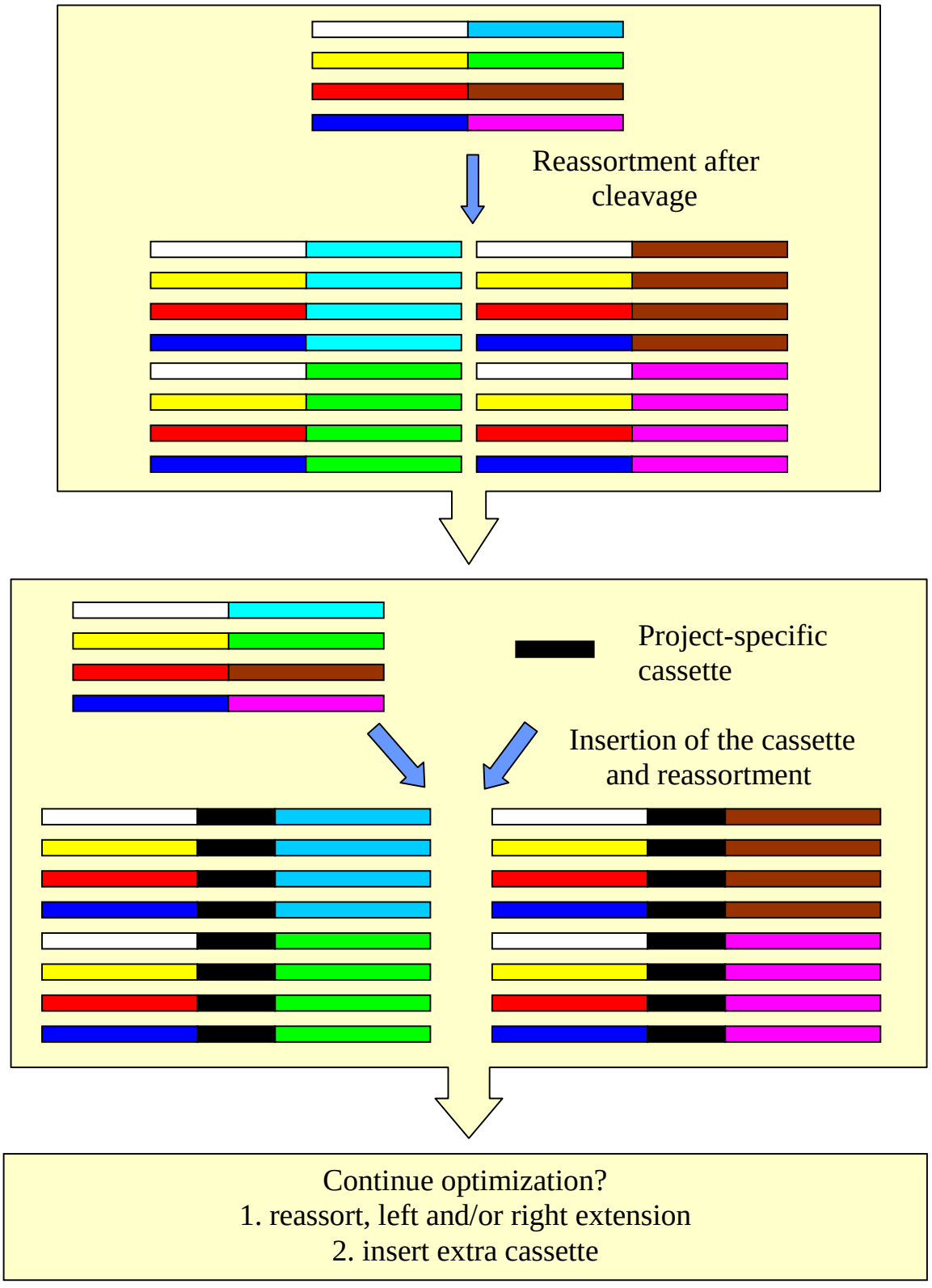


Figure I-8: Summary of the way in which cosmix-plexing® can be applied. The upper section shows the selection of an optimal primary ligand (principle outlined in Fig.I-11). The consensus motif can be inserted as a "project specific" cassette into a secondary "extension" library from which novel variants with optimized flanking sequences can now be isolated. The cosmix-plexing® induced recombination of the flanking sequences can now be used here, once again, to advantage as an evolutionary strategy.

The theoretical advantage which can be obtained through cosmix-plexing® library, can be deduced by a consideration of the procedure of classical clone enrichment during standard panning. “Normal” pannings are performed on a target of interest, and after a few rounds of affinity selection, some enriched variants can be identified. These variants exhibit appropriate properties to interact with the target molecule. Typically, a few clones are enriched, but some unique variants are also present, which are highly related (i.e., show sequence homology), to the enriched dominant clone(s), so that a consensus sequence can be conceived. This consensus sequence represents the essential structure required for the binding to this target. During the first panning round, the future dominant clones are present at extremely low concentration (subnanomolar), as well as many other clones which exhibit weak affinity to the target. As the dominant clone(s) become enriched, during further rounds, compete with the weaker binders which are then eliminated during the washing steps of the panning.

The cosmix-plexing® takes advantage of structural motifs from the low affinity ligands present in early rounds, by recombining them with motifs from the stronger ligands, thus creating novel diversity not present in the initial population. This novel diversity is strongly enriched for the **structural elements required for binding to the target!**

The preselected population recovered after the first round of panning, is the working material which is then submitted to recombination. At this stage, many of the clones present, show an affinity capacity for the target, even if very weak. This preselected population is recombined with a cross-over in the hypervariable region, producing all possible recombinants (in order to produce all possible recombinants (say 10^8) from the preselected population (say 10^4), a particularly efficient protocol had to be designed). By recombining these clones amongst themselves, one reassorts two halves of the hypervariable region in each clone. This raises the chance of bringing together domains which initially alone only allowed weak affinity, but which now exhibit new features which may correspond to a theoretically “perfect” structure for the binding to the target (at least over a region of 13 amino acids, for the size of library used).

Further selection with this optimized library leads to the isolation of new clones which should be optimal regarding to their affinity to the target.

E – Target proteins

The initial aim of this work was to isolate strong ligands to SH3 domains. Aberrant SH3 interactions are sometimes associated with uncontrolled signaling, which is a characteristic of cancer cells (Cowburn and Kuriyan, 1996). SH3 containing proteins have also been shown to be a causative factor in a fusion following chromosomal translocation, leading to tumor (Mayer et al., 1995; Pendergast et al., 1993; Puil et al., 1994; Feller et al., 1998). It was thus decided to develop competitive high affinity ligands to SH3 domains, so as to develop tools for the investigation of the function of SH3 domain containing proteins in the signal transduction cascade and cytoskeleton. These binders might lead to a better understanding of the cells signaling pathway. However, they might also be able to compete natural ligands *in vivo*, and therefore be able to induce or inhibit the SH3 interactions, providing potential therapy agents against cancer, in combination with cell targeting strategies. Considering the high preference of SH3 domains for proline rich ligands, we built biased random peptide libraries containing prolines at three defined positions.

Recently, it has been shown by others, that the EVH1 domains, protein-protein interaction modules implicated in the spatial control of actin assembly, mediate the interaction with their partners, through a proline rich motif. Proteins of the EVH1 domain family have been shown to interact directly with a bacterial surface protein of the intracellular pathogen *Listeria monocytogenes*. Through this interaction, *L. monocytogenes* is able to recruit the host actin filament in order to support its motility, augmenting its virulence. Strong binders to EVH1 domains could disrupt the interaction between *L. monocytogenes* and the EVH1 modules, and therefore, prevent the cell to cell spreading of the bacteria. Since the EVH1 domains share with SH3 domains the property to bind proline rich ligands, the proline biased library generated for the SH3 domains might be highly advantageous to isolate novel EVH1 domain ligands.

The project was therefore extended beyond the initially proposed characterization of high affinity binders to SH3 domains, to include also EVH1 domain studies.

1- SH3 domains

1-1 Function and structure of SH3 domains

The Src Homology 3 (SH3) domains are found in a wide variety of unrelated proteins, many of which are involved in signal transduction, indicating their important role in protein-protein interactions. SH3 domains organize protein complexes within the cell, bring substrates

to enzymes, and regulate enzymatic activities. SH3 domains are involved in pathways that provide potential targets for therapeutic agents.

SH2 and SH3 domains are found together on many proteins (Figure I-9). Evidence suggests that their activities might be coordinated, as shown in the work of Superti-Furga, where mutation of the Src SH3 activates the kinase, and increases the accessibility of the SH2 to exogenous substrates (Superti-Furga et al., 1993).

SH2/SH3 domain containing proteins can be divided into two groups. The first group is formed by enzymes, such as the Src family and phospholipase Cγ. The second group is the adaptor protein group, which do not have any known catalytic activity, such as Grb2, Nck, or Crk (Buday, 1999). The discovery of these adaptors revealed a new mechanism by which signaling proteins can form multiprotein complexes. These complexes are held together by the specific interaction of proline rich motifs with the SH3 domains of the adaptors. The SH2 domains of the adaptor can then recruit the complex to phosphorylated tyrosine residues.

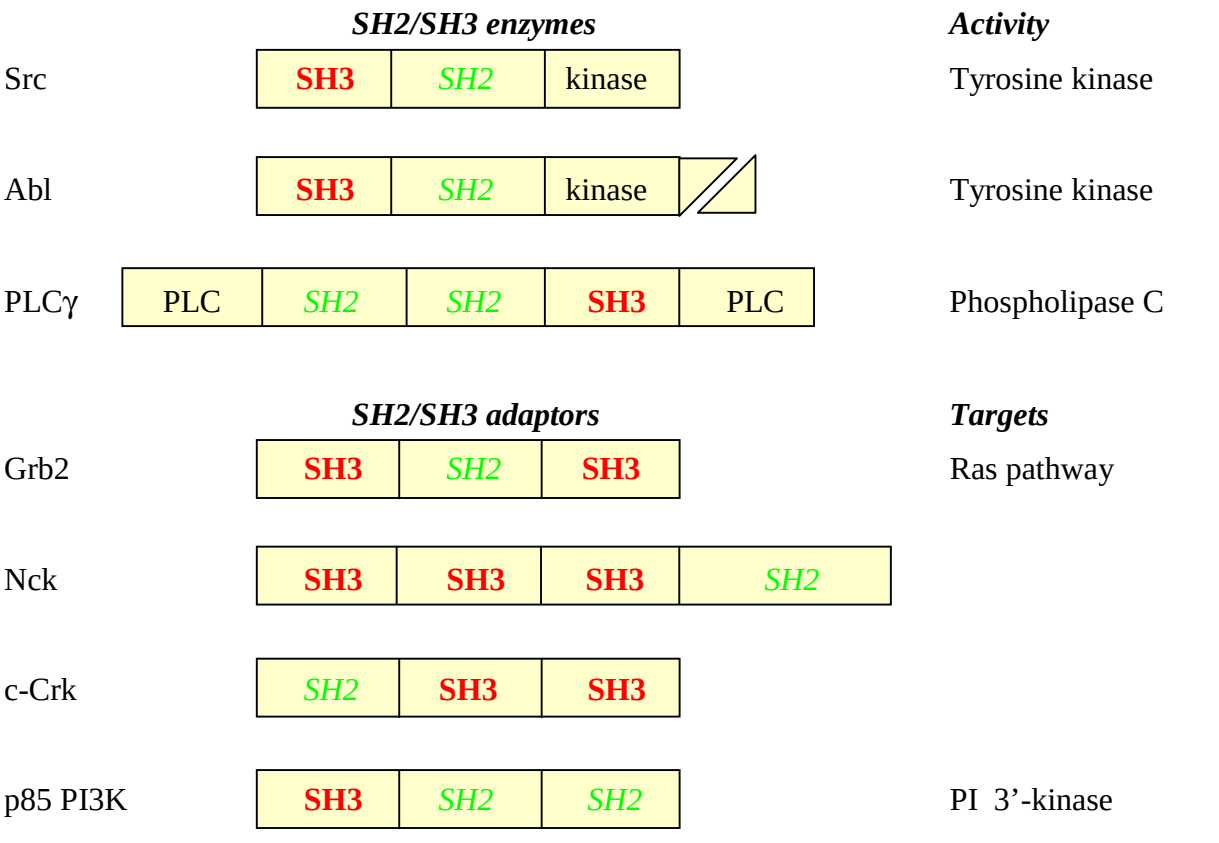


Figure I-9: Structure of selected SH3 containing proteins (Buday, 1999; Pawson and Gish, 1992). The proteins are divided into those with intrinsic enzymatic activity, and those without known catalytic domains, which may act as adaptors, to couple tyrosine kinases to downstream targets.

SH3 domains are protein modules of approximately 60 residues, which can be expressed, purified and crystallized independently of the rest of the protein. The ligand binding pocket is formed by a cluster of hydrophobic residues, and is flanked by two charged and variable loops (n-Src, and RT loops) that contain clustered aromatic residues (Pawson and Gish, 1992; Cohen et al., 1995; Brown and Cooper, 1996).

1-2 SH3 ligands form a left-handed polyproline type II helix

Homologies between natural-, and phage display isolated-ligands, suggest that the binding motifs are 8 to 10 residues long, and are proline rich (Viguera et al., 1994; Lim et al., 1994; Pawson, 1995; Alexandropoulos et al., 1995; Cohen et al., 1995). They present a common PXXP core motif, which interacts with the hydrophobic groove. The specificity for each particular SH3 domain, is determined by the flanking amino acids which allow more extensive contact with a third pocket located in a valley between “n-Src” and “RT” loops. This third pocket is the most structurally diverse in different SH3 domains, and play a major role in binding specificity.

Interactions of proline-rich peptides with SH3 domains takes place through formation of a left handed polyproline type II helix (PPII). This PPII helix contains 3 residues per turn, with a triangular cross-section. One face is contacting the floor of SH3 domain binding groove as shown in Figure I-10 (Cohen et al., 1995; Nguyen et al., 1998).

Adzhubei and Sternberg present interesting results pertaining to PPII helices (Adzhubei and Sternberg, 1993). PPII helices are usually short segments of four to eight residues. However, they are quite voluminous. A PPII helix is two times longer than an α-helix, containing the same number of residues (the length of a PPII segment that includes four C^{α} atoms is approximately equal to the length of a seven residue α-helix).

The most preferred amino acid within PPII helices, is proline. Its advantage is that, due the preferred folding angle at proline residues, proline rich sequences adopt a PPII conformation without a significant loss in conformational entropy. However it seems that proline is not essential within the PPII helices, since 28 of the 96 PPII helices listed by Adzhubei and Sternberg, are formed without a single proline (Adzhubei and Sternberg, 1993). The work of Pisabarro and Serrano also tends to minimize the role of proline in PPII helices. They indicate that prolines within the PPII helices are not essential. They show that proline can be replaced by other amino acids, as long as a replacement and/or a compensating

mutation is introduced elsewhere. It is thus concluded that there seems to be no essential requirement for a particular Pro residue for an efficient interaction with an SH3 domain. However, they could not prove that it was possible to mutate several Pro residues at the same time (Pisabarro and Serrano, 1996). Whatever, the repertoire of SH3 binding peptides might be much larger than previously thought.

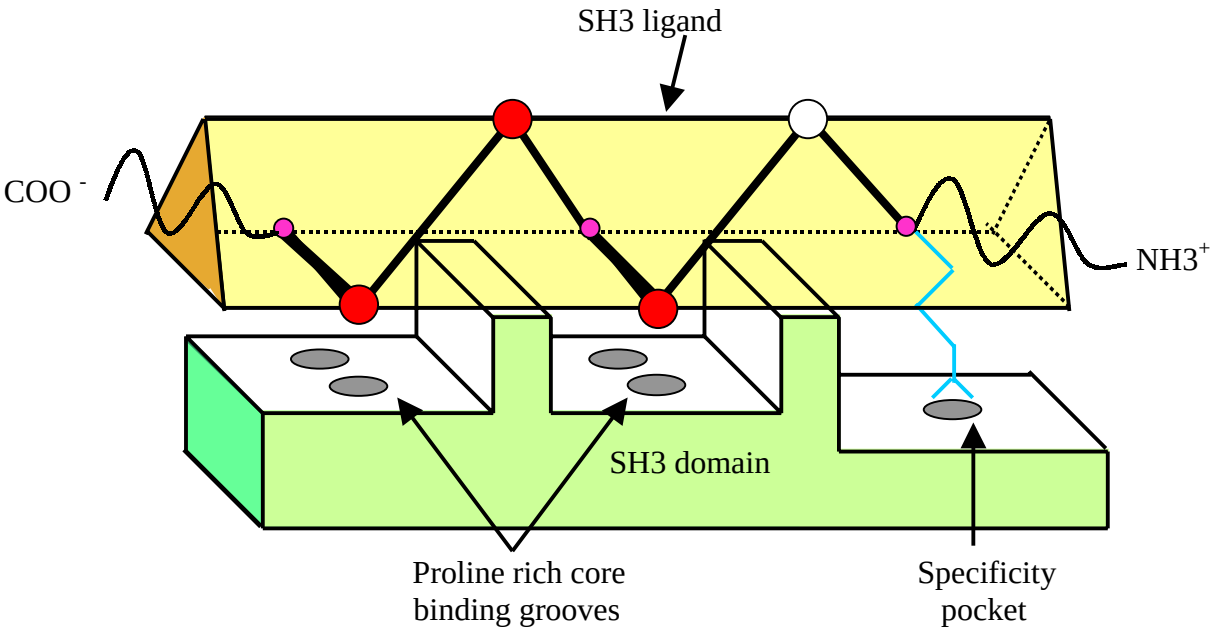


Figure I-10: Peptide/SH3 domain complex interface shown schematically: polyproline type II helix conformation (diagram according to Nguyen et al., 1998). The PPII helix contains 3 residues per turn, with a triangular cross-section, with 3 spines. One face is contacting the floor of SH3 domain binding groove. The residues located on the upper spine cannot make any contact with the SH3 domain. Its role is to stabilize the PPII helix. The red circles indicate positions were proline are advantageous. The blue line between the last residue of the helix, on the N terminal side, and the SH3 domain specificity pocket, indicates a salt bridge.

Other residues commonly present in PPII helices (according to the Chou-Fasman algorithm, i.e. frequencies higher than average in the sample), are Gln, Ser, Arg, Ala and Val (Chou and Fasman, 1978). Interestingly, Chou-Fasman predictions consider Gly as a residue of choice for PPII helices, however, it is not commonly observed. Most of the PPII helices do not form regular interchain hydrogen bonds. Thus, it is assumed that Gly would destabilize the helix.

PPII helices lie mostly on the protein surface. Hydrophilic residues tend to occur commonly in this structure. High accessibility, and orientation of NH and CO groups in PPII helices suggest that potentially the formation of main-chain hydrogen bonds is favorable for

this structure. It is assumed that the most probable partners for this hydrogen bonding are water molecules of the protein hydration shell, as well as water molecules incorporated in protein structure. PPII helices, on the protein surface, form hydrogen bonds with water. Hydration has a strong stability effect on the PPII helices.

1-3 Two classes of SH3 ligands

SH3 domains were considered to be unique in that they can bind their ligands in opposite orientations, being classified as class I and class II ligands, both found in nature (Feng, et al., 1994; Yu, et al., 1994). It has been recently shown that they share this particularity with profilin, which also interacts with highly proline rich motif (Mahoney et al., 1999). Class I and II ligands bind in NH2 - COOH and in COOH - NH2 orientation respectively. The class I ligands for several SH3 domains, show the consensus sequence RXLPPXP, while for the class II ligands the consensus sequence XPPLPXR has been deduced. The major difference between the two classes, is the position of a critical arginine, which forms a salt bridge with a conserved acidic residue in the SH3 domain, and thus, determines the orientation of the ligand. This orientation-determining salt bridge was found to be an important specificity determinant for cellular proteins binding to SH3 domains.

In summary, previous to this study, it was considered that SH3 ligands must contain a central proline core motif, which adopts the PPII helix conformation, and interacts with the aromatic grooves of SH3 domains. This core motif is flanked by N and C terminal amino acids, one of each being assumed to be responsible for the specificity of each SH3 domain. The binding of the flank appears to require the presence of the core, since a peptide having only the flank sequence does not bind the SH3 domain. The core serves as a common anchor used by SH3-binding peptides for initial docking, while the flank allows more extensive contact at the intermolecular interface, thereby providing increased affinity and specificity (Feng, et al., 1995).

1-4 Src SH3 domain

Over the 20 last years, much has been learned about Src and the related Src-family kinases. Src family kinases are evolutionary conserved, with members found in both vertebrates and non-vertebrates. In the mouse, 8 members of the family have been identified: Src, Fyn, Yes, Lyn, Hck, Fgr, Blk, Lck), and one additional has been characterized in the chicken, Yrk. All of these kinases share a common structure: a short amino-terminal membrane anchor, a unique domain characteristic of each individual kinase, the SH3 domain,

a SH2 domain, which binds phosphotyrosine-containing peptide sequences, a catalytic domain which encodes kinase activity, and a short carboxy-terminal tail containing the major regulatory tyrosine residue. In all regions, except the unique domain, all family members show extensive sequence homology (Lowell and Soriano, 1996; Brown and Cooper, 1996).

Src and the other Src family kinases, are believed to associate with the cytoplasmic tails of transmembrane receptors that lack kinase domain, and thus, serve to transduce signals from these receptors. A great deal of redundancy between members of the family is observed. Ligation of receptors for growth factors (PDGF, EGF), cytokines (IL-2, IL-3, IL-6), antigens, antibodies (thrombin, chemoattractants), and adhesion receptors (integrins) have all been shown to result in activation in one or more of the Src family kinase members, in many different cell types (Erpel and Courtneidge, 1995).

Src is found nearly ubiquitous, but highest in brain, platelets, and osteoclasts. Src have essential functions in regulating cell growth and differentiation. However, most cells developed normally in mice with a homozygous disruption in the *src* gene (Soriano et al., 1991). Even cells in which Src is normally expressed at high levels, such as blood platelets, seem to be unaffected, probably due to a compensation by other members of the Src family.

One cell type is affected by the deletion of the *src* gene. In Src deficient mice, osteoclasts are unable to resorb bone effectively, resulting in osteopetrosis (Soriano et al., 1991), suggesting that Src has an essential function in osteoclasts. A hypothesis to explain that on this case, no compensation appears when Src is deficient, is that the other Src family members are not stimulated by the signal recognized by Src, or, more presumably, that Src have unique substrates which cannot be phosphorylated by the other Src family members.

1-5 Nck SH3 domain

The human Nck gene has been localized to the 3q21 locus of chromosome 3 (Huebner et al., 1994). Aberrant genetic rearrangements within this region of chromosome 3 are associated with various human hematopoietic malignancies, suggesting a potential involvement of Nck in oncogenesis (Mitelman et al., 1991). Nck is a 47 kDa protein ubiquitously expressed in mammalian cells (Lehmann et al., 1990), and is composed of three SH3 domains, followed by a single SH2 domain. Nck has no intrinsic catalytic activity and acts as an adaptor molecule, to couple upstream signals, usually those initiated by activation of receptor tyrosine kinases, to downstream signal transducer molecules (McCarty, 1998; Buday, 1999).

Ligand binding to receptor tyrosine kinases induces the receptor chains to dimerization and to transphosphorylate on specific tyrosine residues that provide docking sites for SH2 domains. Nck interacts via its SH2 domain with phosphotyrosine residues in receptor tyrosine kinases (such as the receptors for epidermal growth factor (EGF), platelet-derived growth factor (PDGF), or vascular endothelial cell growth factor (VEGF)). Nck interacts via its SH3 domains with effector molecules containing proline-rich sequences, bringing them to the proximity of ligands-activated receptor tyrosine kinases.

1-6 Nephrocystin SH3 domain

The nephrocystin SH3 domain has been discovered very recently, in the group of Hildebrandt (Hildebrandt et al., 1997a; Hildebrandt, 1998). The nephrocystin gene is contained within a chromosomal region which is deleted in 80% of families suffering from nephronophthisis.

Juvenile nephronophthisis (NPH), an autosomal recessive cystic kidney disease, is the primary genetic cause of chronic renal failure in the first two decades of life. The end-stage renal failure is reached at an average age of 13 years, and leads to death unless treated with dialysis or renal transplantation. The earliest anomaly is the presence of groups of atrophic tubules with an irregularly thickened tubular basement membrane and focal interstitial fibrosis. Diffuse tubulo-interstitial changes and medullary cysts are found later. These early tubular basement membrane abnormalities and the inability to label them with certain anti-tubular-basement-membrane antibodies (Cohen and Hoyer, 1986), suggest that the fundamental defect in nephronophthisis might be the production of an abnormal tubular basement membrane. A number of heterogeneous extrarenal abnormalities have been described, associated with the renal disease (Kleinknecht and Habib, 1992).

A gene locus for NPH, termed *NPH1*, has been located on chromosome 2q12-q13 (Antignac et al., 1993). The *NPH1* genetic region was cloned, as a means to identify the responsible gene, and the presence of a large homozygous deletions of the NPH1 critical region was demonstrated in 80% of NPH1 families (Hildebrandt et al., 1997b). A novel gene, termed *NPHP1*, was characterized, which consists of 20 exons. Almost the entire gene was found to be deleted in 16 of 22 families with NPH1, with one family exhibiting a smaller deletion. It has been shown that there is at least one additional gene locus present in the human genome (Medhioub et al., 1994).

Northern blot analysis of NPHP1 revealed a transcript of 4,5 kb, with strong expression in skeletal muscle, and weak expression in kidney, pancreas and heart. The fact

that NPHP1 is expressed only weakly in kidney should not diminish its candidacy as the disease gene, as there are now many instances, where tissue expression levels of disease genes do not correlate with the site of clinical manifestation (Meindl et al., 1996). The sequence data base comparison of the full length cDNA sequence showed that *NPHP1* gene is unrelated to known gene families. However, a short fragment of the deduced amino acid sequence of its gene product, termed nephrocystin, shows strong similarity to SH3 domains, particularly to the *c-crk* proto-oncogenes from mouse and man, as well as to other SH3 containing genes of *Drosophila*, *Caenorhabditis elegans*, and *Schizosaccharomyces pombe*. The potential function of the SH3 domain in the *NPHP1* gene product is unknown.

2- EVH1 domains

The Ena/VASP homology (EVH1) domain is a protein interaction module found in several proteins that are involved in transducing migratory and morphological signals into cytoskeletal reorganization (Prehoda et al., 1999). EVH1 specifically recognizes proline-rich sequences in its binding partners and directs the localization and formation of multicomponent assemblies involved in actin-biased motile processes (Mitchison and Cramer, 1996; Laurent et al., 1999) and neural development (Tu et al., 1999).

The EVH1 domains composed family of adapter proteins, including **Enabled** (Ena), identified in genetic screens for dominant mutations that alleviate phenotypes associated with mutations in the *Drosophila* homologue of Abelson tyrosine kinase (Abl), **Vasodilator-stimulated Phosphoprotein** (VASP), characterized as a major substrate for cAMP- and cGMP-dependent protein kinases in human platelets, two murine proteins, **Mena** and **Evl,** identified by their similarity to Ena, the **Wiskott-Aldrich syndrome protein** (WASP), or **Vesl/Homer**, which belong to a family of postsynaptic proteins that are thought to function in synaptic plasticity (Brakeman et al., 1997; Kato et al., 1997).

The proteins share a similar overall structure: the highly conserved N-terminal Ena-VASP Homology Domain 1 (EVH1), a central proline rich sequence, containing SH3 or profilin binding motifs (except for Vesl/Homer), and the conserved C-terminal EVH2 domain (or WASP homology 2, WH2 for WASP).

Specific interaction between EVH1 domains and their ligands serves to mediate the proper localization of proteins from the Ena/VASP family and to link them to other molecules with which they cooperate. EVH1 domains localization occurs through binding to proline-rich motifs of consensus sequence FPPPP (Niebuhr et al., 1997). Subsequently, they act as scaffolds to recruit or activate downstream components of the actin assembly machinery.

FPPPP motifs are found in several proteins involved in transmitting external signals to the actin cytoskeleton, including the focal adhesion proteins zyxin and vinculin, and the axon guidance receptors SAX-3/Robo.

The proline rich EVH1 ligands, as for SH3 domains, bind to the EVH1 domain through a polyproline type II helix conformation (Prehoda et al., 1999).

The cellular proteins VASP and Mena, have been shown to interact directly with a bacterial surface protein of *Listeria monocytogenes*: ActA. This interaction is mediated by the EVH1 domain which bind to a minimal consensus binding sequence E/DFPPPPTD/E, a motif repeated four times within the bacterial ActA protein (Niebuhr et al., 1997).

L. monocytogenes is a widespread, rapidly growing, Gram-positive bacterium, capable of invading a variety of host cells, that causes food-borne infections in animals and humans with severe implications, especially for pregnant women, newborns, and immunocompromised individuals. Its virulence and cell to cell spreading is dependent on its ability to exploit the actin-based cytoskeleton of host cells to support its motility. ActA is necessary and sufficient for the recruitment of host actin filament. It contains two separable domains required for interaction with the actin cytoskeleton: a N-terminal 23 amino acid region, essential for actin filament nucleation, and a proline-rich repeat region, required for efficient actin filament recruitment.

ActA initiate a cascade of events leading to formation of a specific F-actin structure, and subsequent actin-based movement in favor of *L. monocytogenes*, through its interaction with EVH1 domains. This suggests that VASP, Mena and their relatives act as efficient accelerators of actin filament assembly, probably by delivering profilactin to sites of high actin filament dynamics (Niebuhr et al., 1997). ActA attracts profilactin to the interface of the bacterium and its actin tail by recruiting VASP and Mena, thereby enhancing local actin polymerization and enabling rapid actin-based movement. VASP and Mena are therefore required for efficient intracellular motility and virulence of *L. monocytogenes* (Niebuhr et al., 1997; Carl et al., 1999; Laurent et al., 1999; Prehoda et al., 1999). Furthermore, EVH1 domains bind strongly to ActA than to natural ligands. Therefore, the study of EVH1 interactions is of medical interest, with respect to *L. monocytogenes* pathogenesis.

3- The use of fusion proteins for higher expression: GST fusion proteins

During the early stage of recombinant DNA technology, it was thought that a strong promoter and a start codon at the beginning of the gene would be sufficient for good

expression in *Escherichia coli*. Since then, it has been learned that in addition to a promoter and a start codon, good expression requires that the mRNA contain a ribosome-binding site that is not blocked by mRNA secondary structure. The level of expression is also affected by codon preferences, especially in the second codon of the gene (Stormo et al., 1982), and may be affected by the coding sequence in other ways that are not yet well understood. These problems can be solved by altering the sequence preceding the start codon and/or by making changes in the 5' end of the coding sequence that do not change the protein sequence, taking advantage of the genetic code.

However, since much of this data on translation variables is anecdotal, it is often quicker to solve these problems by making fusions between genes. In this approach, the cloned gene is introduced into an expression vector fused to a sequence (carrier sequence) coding for a highly expressed protein (carrier protein). The carrier sequence can be from any gene that is strongly expressed in *E. coli.* The carrier sequence provides the necessary signals for good expression.

These carrier regions can often be exploited in purifying the protein, either with antibodies, or with affinity purification specific for the carrier protein. Alternatively, unique physical properties of the carrier protein can be exploited to allow selective purification of the fusion protein.

There are two problems often encountered when expressing fusion proteins. The first one concerns the solubility of the fusion proteins. In many cases, a high level protein expression leads to the formation of inclusion bodies (very dense aggregates of insoluble protein and RNA that contain most of the expressed protein). This difficulty can be solved by lowering the expression level of the protein, or by using another strain to propagate the fusion proteins (La Vallie et al., 1993; Schein, 1989).

The second problem deals with the stability of the expressed protein and the presence of the carrier protein. Fusion proteins are sometimes cleaved *in vivo* at the junction/linker of the fusion protein, which creates problems if the carrier protein is to be used as an aid in purification. It might happen that the carrier domain and/or the expressed protein do not fold correctly and might be degraded by *E. coli* proteases. One solution to stabilizing the fusion proteins is to express them as insoluble aggregates. Another possibility is to use *E. coli* strains deficient in known proteases. In addition, shorter induction times may increase the ratio of full-length proteins to breakdown products, although the total amount of protein will be lower.

In some cases it is advantageous to remove the carrier protein in order to facilitate biochemical and functional analyses. In most cases, the linker between the carrier protein, and the protein to be expressed, contains cleavage possibilities, either for chemical cleavage (cyanogen bromide or low pH which allow cutting respectively after Met or between Asp and Pro, as examples) or for enzymatic cleavage (e.g. factor Xa, thrombin, enterokinase, renin or collagenase). All of these enzymes have extended substrate recognition sequences which greatly reduces the likelihood of unwanted cleavages elsewhere in the protein, although cleavage is commonly inefficient.

The fusion protein can be coupled to many different carrier proteins, such as β-galactosidase, *trp*E, maltose-binding protein (MBP), thioredoxin (Trx), or glutathione-S-transferase (GST). The GST fusion protein is the only fusion expression system discussed in this work as it is the system applied in our laboratory. The main advantage of GST fusion proteins over these other methods is its use during purification, and at a later stage, in its use for immobilizing the target during panning and measuring kinetic binding constants. Indeed, for β-galactosidase, *trp*E, MBP and Trx, the denaturing reagents used during purification can be expected to alter the antigenicity and functional activity of the purified product. However, GST can be purified by affinity chromatography on immobilized glutathione followed by competitive-elution with excess reduced glutathione (Simons and Vander Jagt, 1977).

pGEX vectors have been designed to express foreign polypeptides as fusions with GST. Such fusion proteins are soluble and are easily purified from lysed cells under nondenaturing conditions.

Each pGEX vector contains an open reading frame encoding GST, followed by unique restriction sites, followed in turn by termination codons in all three reading frames. The GST carrier of pGEX vectors can be eventually removed by site-specific proteolysis, or by chemical hydrolysis at low pH.

Expression of the GST fusion proteins is controlled from a p_{tac} promoter. The *lac* repressor (product of the *lacI* gene) binds to the promoter, repressing the expression of the fusion protein. Upon induction with IPTG, derepression occurs and GST fusion protein is expressed. Typical yields of fusion protein vary from 1 to 3 mg/liter up to 10 mg/liter.

F – BIAcore – real time interaction measurements

The BIAcore is a biosensor-based analytical system, designed for functional characterization of protein-protein, protein-DNA, and ligand-receptor interactions in real time. As all the interactions are dynamic processes, the ability to measure precisely their kinetic parameters and determine how fast one molecule binds to and dissociates from another is very important for understanding the regulation of the biological processes involved.

The BIAcore is based on a surface plasmon resonance (SPR) to monitor the adsorption of biomolecules on a sensor chip. This optical technique measures changes in refractive index of the medium very close to the surface. Such changes are directly proportional to the amount and molecular weight of the macromolecules bound at any time.

The system includes a sensor chip to which the ligand can be immobilized on a matrix composed of flexible dextran, a miniaturized fluidics cartridge for the transport of analytes and reagents to the sensor chip, a SPR detector (Jönsson and Malmqvist, 1992; Malmqvist, 1993). A laser source emits a plane-polarized light which is aimed through a prism, on a gold covered sensor chip where the targets of interest are immobilized. When the analyte binds to this target, the beam angle is altered, and monitored, so that the intensity of the interaction can be determined. The measurements are performed continuously and the angle differences are expressed in arbitrary resonance units, so that a real time plot (sensorgram) of the sensor response, in resonance units, versus time can be obtained (Panayotou et al., 1993).

In this work, the BIAcore was used to study the binding kinetics of peptides to our target molecules. The adsorption rate can be measured using BIAcore, since the adsorption of an analyte to a surface-immobilized ligand is carried out at a constant flow rate whereby the concentration of the analyte is kept constant. This adsorption can be regarded as a pseudo first order reaction. From this, the association rate constants can be obtained by carrying out a series of adsorption experiments with different analyte concentrations. The decrease in complex concentration on the surface after the sample pulse has passed can be used for the determination of dissociation rate constants.

The BIAcore has been shown to be valuable in several other applications, such as determination of interacting ligands (Malmborg et al., 1996), protein/DNA interactions (Sibille et al., 1997), or epitope mapping (Fägerstam et al., 1990). In addition, some groups used the BIAcore as an alternative affinity selection procedure, to select binders from phage displayed antibody libraries. Malmborg et al. injected Fab phage stocks displaying antibodies specific for hen egg lysozyme, over a biosensor chip containing immobilized lysozyme. The bound phage were eluted, and characterized as antigen-specific (Malmborg et al., 1996).

G – Testing the functionality of the selected peptides, *in vivo*

Structure-function analysis of signal transduction proteins, can be aided by synthetic peptide analogs, representing putative sites of intracellular protein-protein interactions. Unfortunately, the cell membrane is usually impermeable to synthetic peptides, or antibodies. Therefore, a wide variety of methods have been proposed for the delivery of proteins and other macromolecules into living cells for either experimental or therapeutic uses. These include microinjection of antibodies and peptides to individual cells (Diacumakos, 1978; Wang et al., 1982); the use of membrane disrupting pore-forming reagents (Wang et al., 1982); electroporation (Stauffer et al., 1997); liposome encapsidation (Renneisen et al., 1990; Wu et al., 1991); bacterial toxins (Prior et al., 1992; Stenmark et al., 1991); or receptor-mediated endocytosis (Ishihara et al., 1990; Wilson et al, 1992; Leamon and Low, 1992). Most of these methods are either inefficient or time-consuming, cause appreciable cell death, or result in uptake into intracellular vesicles without efficient cytoplasmic delivery. Several approaches rely on binding of macromolecules to the cell surface, followed by internalization via the endocytic route. However, since proteins that have entered this pathway remain enclosed within the lipid vesicles, they do not have access to the cell cytoplasm. Alternatively, transfection experiments can be used to introduce DNA encoding truncated or mutated intracellular proteins (Henkel et al, 1992; Perlmutter et al., 1996). Such approaches have, however, severe limitations which obstruct progress in analysis of intracellular proteins.

During the past years, new strategies to deliver functional peptides to cells, have been developed, to probe intracellular protein-protein interactions. Techniques used to transfect proteins or peptides within cells are briefly described and compared in this chapter.

1- Conventional methods for transfection

1-1 *Antennapedia* homeodomain peptide

The 60 residue homeodomain of the *Drosophila Antennapedia* has the ability to rapidly cross the plasma membrane of neuronal cells, and accumulate in their nuclei (Joliot et al., 1991a). The third helix within the homeodomain, encompassing 16 residues, was shown to translocate through the cell membrane in a temperature-, energy-, and receptor-independent manner (Derossi et al., 1996). This discover gave rise to its application as a nuclear targeting agent. However, the internalization of *Antennapedia* homeobox domain requires a neuronal cell adhesion molecule decorated with α-2, 8-polysialic acid. Therefore, this domain has special predilection for neuronal cells, since they express this neuronal cell adhesion

molecule. To some extent, other cells (such as fibroblasts: Joliot et al., 1991b) express this adhesion molecule, and thus, allow uptake at a low level.

However, this system cannot be commonly applied, as it is restricted to particular types of cells, and presents interest only if imported peptides are to be localized in the nucleus.

1-2 Myristoylated peptide

Another approach to cellular import of peptides utilizes modification of peptides by attachment of a fatty acid as a means to overcome the permeability barrier of the plasma membrane. Amino-terminal myristoylated peptides are cell permeable (Eichholtz et al., 1993). Attachment of a myristoyl group to the amino-terminus of the peptide, provides a suitable membrane anchor, and therefore, it potentially restricts the use of this imported peptide to the cytoplasmic face of the plasma membrane.

Furthermore, these complex molecules exhibit intolerable cytotoxicity, limiting their use for structure-function analysis of intracellular protein in living cells.

1-3 Signal sequence h region

Cell-permeable peptide import (CPPI), based on the signal sequence hydrophobic region (h region), takes advantage of its ability to translocate across cellular membranes, due to a short membrane translocating sequence (MTS) (Lin et al., 1995; Hawiger, 1997). This region is present in Kaposi fibroblast growth factor (FGF), and in integrin β_3 signal sequences (16 and 15 residues respectively). The ability of signal peptides to insert into membranes, and their *in vivo* function, correlate with the residue-average hydrophobicity of their hydrophobic cores. The short h region of a signal peptide can be used as a carrier for import of a cargo peptide. The h region can be placed at either the N- or C-terminus. Rojas et al. (1998), succeeded to import a 41 kDa GST-Grb2SH2-MTS protein into NIH 3T3 cells.

When such cell-permeable peptides are associated with a short peptide sequence called the nuclear localization signal (NLS), the imported peptide is found to be present in the nucleus. Typically, NLS motifs represent a cluster of basic residues (arginine and lysine) flanked by acidic amino acids and/or prolines. The NLS is recognized, and binds to the cytoplasmic face of a NLS receptor, on the nuclear pore, with the help of an accessory protein (p90), and the NLS bearing protein is shuttled to the nucleus.

The intracellular localization of imported C-terminal MPS fusion peptides, reaches a maximum concentration within 45 minutes, at 37°C, and is detectable in cells, up to 180

minutes after peptide addition. The intracellular concentration of imported peptides reaches 4% of the peptide added to the media. Therefore, this technique appears to be limited by the short life time of the imported peptide within the cells, and a relatively low efficiency, since 96% of it, will not be transfected. In addition, when presence of imported peptides is required within the nucleus, a more complex system has to be faced: the presence of the NLS is required in the construct, but moreover, the translocation involves other proteins, and depends on the p90 accessory protein (see Powers and Forbes, 1994, for a minireview on nuclear import).

2- HIV-1 Tat protein

In 1988, Green and Frankel independently discovered that human immunodeficiency virus type 1 (HIV-1) Tat protein was able to cross cell membranes. When, in 1994, Fawell et al. demonstrated that chemically cross-linking Tat to heterologous proteins, conferred the ability to transduce into cells, it has been submitted to criticism, as it was thought that the Tat protein was highly toxic for the cells. Indeed, it has been shown that Tat effects include alteration of normal neuronal and astrocyte organization (Kolson et al., 1993), upregulation of β_2 integrins in monocytes, thereby increasing their adhesiveness (Lafrenie et al., 1996), angiogenic activities (Albini et al., 1995), and apoptosis of cultured human blood cellular mononuclear cells (Li et al., 1995).

However, these toxicity difficulties have been overcome by using truncated Tat domains for the fusion with the heterologous proteins. Tat is an 86-amino acid protein, involved in the replication of HIV-1. A region of the Tat protein, extended from residues 37 to 72, and centered on a cluster of basic amino acids allows the internalization of fused proteins. Originally, truncated Tat (37-72) was selected because it retains the basic domain implicated in uptake, and lacks the cysteine-rich region of Tat (residues 22 to 37) that has complicated the handling and analysis of fusion-peptides (Fawell et el., 1994). Vivès et al. (1997) showed that, unexpectedly, the domain extending from amino acids 37 to 47, and which adopts an α-helical structure, with amphipathic characteristics, is not required for efficient and fast cell uptake. It even appears that a high concentration of peptides with full length α-helix decreased the cell viability. The Tat segment essential for Tat translocation and nucleolar localization, was delineated as the previously described cluster of basic amino acids (Arg-Lys-Lys-Arg-Arg-Gln-Arg-Arg-Arg, from positions 49 to 58), which additionally contains a nuclear localization signal (NLS) sequence. NLS causes the ever-open central

channel of the pore (90 Å) to expand transiently up to 260 Å in diameter to allow nuclear protein entry (Powers et al., 1994).

The mechanism by which Tat traverses a membrane remains unclear. It is not believed any longer that endocytosis would be involved, as low temperature incubation do not alter cellular uptake. It has also been proved that a significant membrane disruption by Tat basic peptides, or a well defined internalization pathway were not involved in the import (Vivès et al., 1997).

It has been demonstrated that Tat binds to specific cell surface proteins, raising the possibility that delivery is receptor-mediated. However, since delivery at present appears to be independent of cell type, then, if such a receptor exists, it would have to be ubiquitous.

Currently, the most likely hypothesis for the Tat translocation mechanism, involves a tight ionic interaction between the basic groups of the peptide side chain, and the negative charges of the phospholipid heads, which would induce a local invagination of the plasma membrane. The local reorganization of the phospholipid bilayer would then lead to the formation of inverted micelles with the peptide enclosed in the hydrophilic cavity and ultimately to the cytoplasmic release of the peptide. Since protein transduction is thought to target the lipid bilayer component of the cell membrane, in principle, all mammalian cell types should be susceptible to import of proteins.

The short cluster of basic amino acids of Tat protein (49-58) has been extensively used in Dowdy's laboratory, where a new strategy for the transfection has been developed. In the work of Fawell, or Vivès, the heterologous protein was chemically cross-linked to the Tat fragment. However, the synthesis of peptides carrying clusters of basic residues (and succession of arginine in particular) often lead to coupling problems. In addition, it appears that optimal uptake of Tat required the presence of the lysomotropic agent chloroquine to reduce protein degradation.

In Dowdy's approach, rather than depending on the uncertainties and inefficiencies of chemically cross-linking proteins, a bacterial expression vector (pTAT-HA) was constructed, to produce in-frame Tat fusion proteins. Because of reduced structural constraints, high entropy, denatured proteins may transduce more efficiently into cells than low entropy, correctly folded proteins. Once inside the cell, transduced denatured proteins may be correctly refolded *in vivo* by host chaperones (such as HSP90). Therefore, the proteins are expressed and purified in a urea-denaturating protein purification protocol (Nagahara et al., 1998).

Tat fusion-proteins prepared in this way rapidly transduced in 100% of cells, achieving maximum intracellular concentration in less than 10 min, in a concentration-dependent manner. In addition, they retain known biological and biochemical activities. Around 60 Tat fusion- proteins have been successfully transduced in Dowdy's laboratory, in a variety of cell types (e.g. fibroblasts, bone marrow stem cells, osteoclasts, or NIH 3T3 cells). Tat fusion-proteins readily transduced into all cells present in whole blood, including both nucleated and enucleated cells. Confocal microscopy analysis has demonstrated that transduced proteins are localized in both cytoplasmic and nuclear compartments, and not just attached to cellular membranes (Nagahara et al., 1998).

Since they developed this system, Dowdy and coworkers already proved that the method is applicable to delivery of therapeutics. By using a "Trojan horse" strategy, they managed to specifically kill HIV-infected cells (Vocero-Akbani et al., 1999). They engineered a modified caspase 3 protein (when activated, caspase 3 induces apoptosis), fused to the Tat protein. This fusion protein transduced in 100% of infected and uninfected cells. However, because of the substitution of endogenous cleavage sites, for HIV proteolytic cleavage sites, Tat-caspase 3 is only activated by HIV protease, in infected cells, resulting in apoptosis, whereas, in the uninfected cells, it remains in the inactive form. This discovery represents an alternative of considerable advantage in treatment of HIV patients, whereby death or apoptosis is achieved in infected cells, thereby limiting or eliminating virus production, while leaving uninfected cells unharmed.

Recently, they succeeded to deliver a 120 kDa biologically active Tat fusion-protein to all tissues in mice, including the brain, by intraperitoneal injection (Schwarze et al., 1999). Previously, importing of bioactive peptides across the blood-brain barrier, was restricted to small (maximum six amino acids) and highly lipophilic peptides. These results open new possibilities for direct delivery of proteins into patients, in the context of development of vaccines, or protein therapies for cancer and infectious diseases.

This Tat fusion system was chosen to transfect ligands into cells in this work. We considered that these novel ligands, isolated with the cosmix-plexing® approach of phage display, might be strong enough, to compete the natural SH3 ligands *in vivo*. This method gives us the opportunity to confirm it.

H – Aim of the work

Phage display is probably the most powerful of the combinatorial library approaches. However, as observed in recent literature, the real challenge currently, is to generate an artificial evolution, and thus to increase the diversity of the libraries, by the introduction of modifications (mutations or recombinations) within the peptides or proteins displayed on the phage surface. The aim of this accelerated evolution, is to identify a few altered variants with optimized binding capacities for their target. This is also the aim of this work.

To reach this goal, to establish a system allowing the isolation of high affinity binders, SH3 domains were chosen as a model system. SH3 domains are essential to regulate cellular interactions, since they are involved in signal transduction pathways. It was considered that increase of the present knowledge about SH3 domains is of interest in developing therapeutic agents, e.g. against cancer, since aberrant SH3 interactions are sometimes associated with uncontrolled signaling, which is a characteristic of cancer cells (Cowburn and Kuriyan, 1996). Furthermore, as many groups concentrated during the last few years on SH3 domains, much is known about their ligands preferences, giving us also the opportunity to compare results obtained with different phage display approaches, and to determine which of them are really advantageous. Therefore, SH3 domains appear as a model of choice to investigate methods which would allow the isolation of high affinity ligands.

The ligand preferences of the SH3 domains are based on a central proline rich motif. This feature allows the creation of dedicated biased libraries. A second family of proteins, the EVH1 domains, were also investigated during this work, as they also show preference for proline rich domains. Their analysis was a further validation of the strategies chosen.

Two strategies were investigated to isolate and characterize very strong binders against SH3 and EVH1 domains. The first one, commonly used for different kind of targets, deals with the utilization of biased constrained libraries. It was unfortunately unsuccessful. The second approach takes advantage of a DNA recombination strategy developed by John Collins (Collins and Röttgen, patent applied, 1997), the cosmix-plexing® technology. The goal of this part of the work was to test whether or not the use of this technology, combined to phage display methods, is advantageous compared to existing technologies.

A brief overview of the aims and the strategies employed in this work begins with the generation of dedicated libraries to SH3 and EVH1 domains. These libraries might allow the characterization of binders, during affinity selection procedures, against both families of

domains. The next objective was to recombine the clones of the dedicated cosmix library among themselves, and to select, from this secondary bank, ligands which present optimized properties for SH3 and EVH1 domains. It was considered feasible that such ligands would be able to compete *in vivo* with natural ligands of SH3 and EVH1 domains, and thus, open the possibility of elaborating cellular models where the signal transduction pathways could be tightly controlled and regulated. These models could open numerous novel possibilities of developing medical applications.

II - MATERIAL AND METHODS

A- Material

1- Chemicals

All Chemicals used in the context of this thesis were of analytical quality. Aqueous solutions were prepared with double deionized (Milli Q) water. Where applicable solutions were autoclaved.

2- Devices

Electrophoresis

Agarose gel: Horizontal gel electrophoresis system, model H3 and horizon 58. Bethesda research laboratories, Neu Isenberg, Germany.

Polyacrylamide gel: Vertical gel electrophoresis system, model V16. Bethesda research laboratories, Neu Isenberg, Germany.

Power supply: Power Pack P30, Biometra Biomedizinische Analytik GmbH Göttingen, Germany.

PCR reactions

DNA Thermal Cycler. Perkin Elmer GmbH, Weiterstadt, Germany

DNA sequencing

ALF-DNA-Sequencer and ALF-Manager software, version 2,5. Pharmacia Biotech Europe GmbH, Freiburg, Germany.

Applied Biosystems ABI Prism 310, and ABI Data collection software, version 2.1. Perkin Elmer Applied Biosystems GmbH, Weiterstadt, Germany.

Electroporation

Gene pulser and pulse controller (0,2 cm cuvettes). Bio-Rad laboratories GmbH, Munich, Germany.

Centrifuges

Biofuge Fresco. Heraeus Sepatech GmbH, Osterode, Germany.

Sorvall RC5C. Du Pont de Nemours GmbH, Bad Homburg, Germany.

Photometer

Multiskan MCC/340 MKII. Bartholomey Labortechnik, Alfter, Germany.

GeneQuant II RNA/DNA calculator. Pharmacia Biotech Europe GmbH, Freiburg, Germany.

ELISA plate reader

Labsystems Multiscan MS, Labsystems, Frankfurt, Germany.

Affinity studies

BIAcore 2000, and biacontrol software version 3.0. Pharmacia Biosensor, Uppsala, Sweden.

3- Computer software

ALF-Manager software, version 2,5. Pharmacia Biotech Europe GmbH, Freiburg, Germany.

ABI Data collection software, version 2.1. Perkin Elmer Applied Biosystems GmbH, Weiterstadt, Germany.

Vector NTI 5.0. Informax Inc., Gaitherburg, MD 20877, USA.

Biacontrol software version 3.0. Pharmacia Biosensor, Uppsala, Sweden.

4- Oligonucleotides

Oligonucleotides were used for cloning, PCR amplification and DNA sequencing, which were ordered from Gibco BRL life technologies Ltd, Paisley, United Kingdom or MWG (Ebersberg, Germany).

4-1 Generation of the constrained libraries: pSKAN8-HypD, -HypE, -HypF, -HypG

4-1-1 Hypervariable oligonucleotides

pSKAN8-HypD

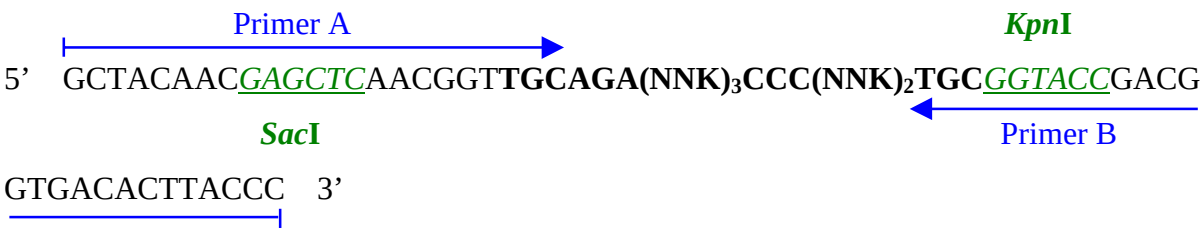
Primer A — ***Kpn*I**

5’ GCTACAAC*GAGCTC*AACGGT**TGC(NNK)$_4$CCANNKAGATGC***GGTACC***GACGGT

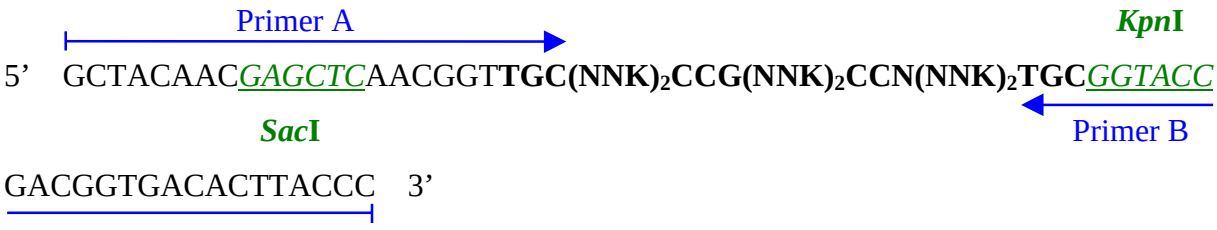
***Sac*I** — Primer B

GACACTTACCC 3’

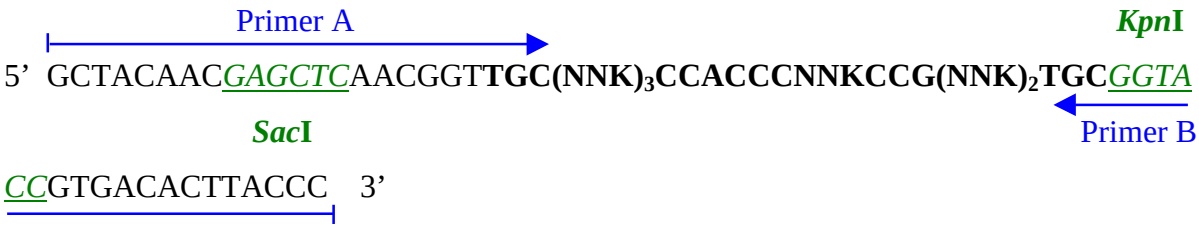
pSKAN8-HypE



pSKAN8-HypF



pSKAN8-HypG



N stands for A, C, G or T

K stands for C or G

4-1-2 PCR primers

Primer A 5‘ biotin GCTACAACGAGCTCAACGGTTGC 3‘

Primer B 5‘ biotin GGGTAAGTGTCACCGTCGGTACCGCA 3‘

4-1-3 Sequencing primer

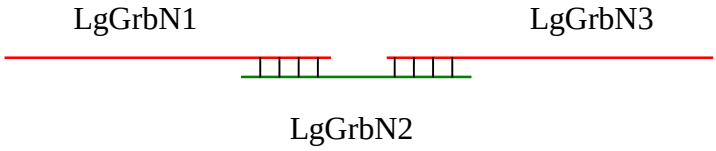
#2897 5’ GGAGGTCTAGATAACGAGG 3’

4-2 Display of a short and a long Grb2N ligand, as pIII fusion peptides

4-2-1 Short Grb2N ligand

4-2-2 Long Grb2N ligand

Due to the size of this DNA fragment, the sequence was divided in three overlapping oligonucleotides (the overlapping regions are represented in italic), leading to the desired product after PCR. The following is representing the overlaps between the three oligonucleotides. Strands 5' to 3' are represented in red, and complementary strands (3' to 5') in green.



LgGrbN1

Primer A

5' CTCAACGGTTGCCCCAAATGGGAC*AGTCTCTTGCCT* 3'

LgGrbN2 (complementary strand)

5' *AAAGCAGGTGGT*AATGC*AGGCAAGAGACT* 3'

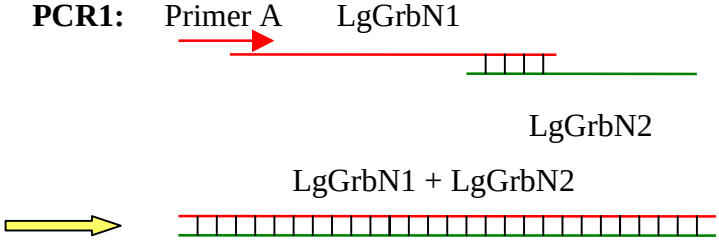
LgGrbN3

5' *ACCACCTGCTTT*CACTGTTGAAGTTTGCGGTACCGAC 3'

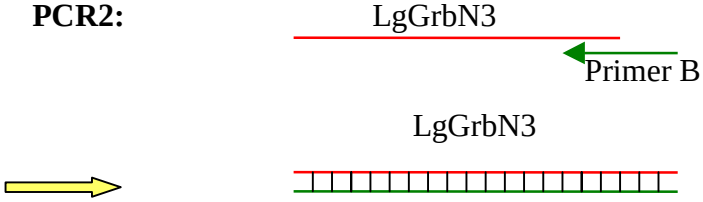
Primer B

The reconstitution of the total fragment was performed in three PCR.

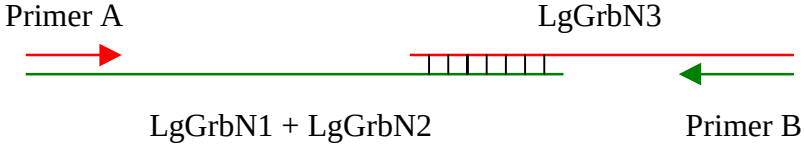
PCR1: Primer A LgGrbN1



PCR2:



PCR3: PCR1 product + PCR2 product



LgGrbN1 + LgGrbN2 + LgGrbN3

4-2-3 PCR Primers

Primer A 5‘ biotin GCTACAACGAGCTCAACGGTTGC 3‘

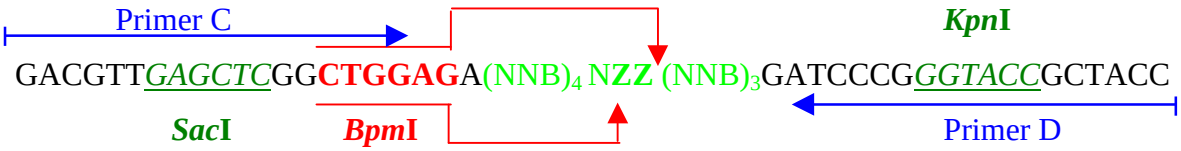
Primer B 5‘ biotin GGGTAAGTGTCACCGTCGGTACCGCA 3‘

4-2-4 Sequencing primer

#2897 5’ GGAGGTCTAGATAACGAGG 3’

4-3 Generation of the CPLPPXP library: 8 amino acid primary library

4-3-1 Hypervariable oligonucleotide



N stands for A, C, G or T

B stands for C, G or T

ZZ represents AM (AA, AC), CT, GG, TS (TC, TG)

4-3-2 PCR primers

Primer C 5’ biotin GAC GTT GAG CTC GGC TGG 3’

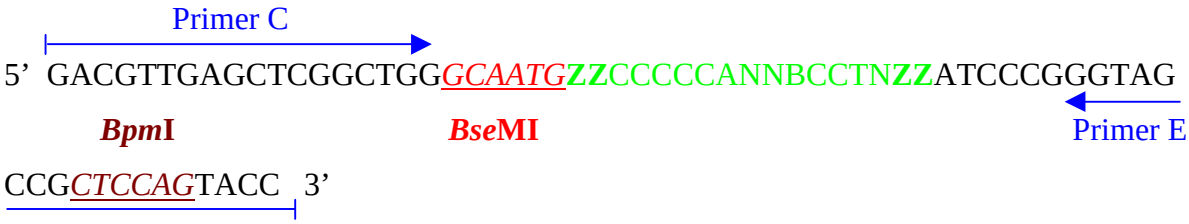
Primer D 5’ biotin GGT AGC GGT ACC CGG GAT 3’

4-3-3 Sequencing primer

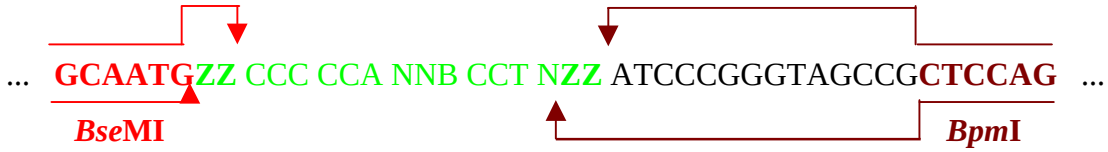
#2897 5’ GGAGGTCTAGATAACGAGG 3’

4-4 Generation of the CPLPPXP library: project specific proline rich cassette

4-4-1 Hypervariable oligonucleotide for the proline rich cassette



Schematic restriction map of *Bpm*I and *BseM*I type IIs restriction enzyme:



N stands for A, C, G or T

B stands for C, G or T

ZZ represents AM (AA, AC), CT, GG, TS (TC, TG)

4-4-2 PCR primers

Primer C 5’ biotin GAC GTT GAG CTC GGC TGG 3’

Primer E 5’ biotin GGT ACT GGA GCG GCT ACC 3’

4-4-3 Sequencing primer

#2897 5’ GGAGGTCTAGATAACGAGG 3’

4-5 Expression of GST fusion nephrocystin SH3 domain

4-5-1 PCR primers

***Sma*I**

Nephro1: 5’ CGTCG*CCCGGG*AATTCCACAAATGG 3’

***Sal*I**

Nephro2: 5’ TAGCTTGGCTGCAG*GTCGAC*GGATC 3’

4-5-2 pGEX 5’ sequencing primer (Pharmacia Biotech, Freiburg, Germany)

5’ GGGCTGGCAAGCCACGTTTGGTG 3’

4-6 Expression of Tat fusion peptides

4-6-1 PCR primers

***Xho*I**

TATfwd: 5’ TATATTC*CTCGAG*CTCGGCTGGAGA 3’

***Eco*RI**

TATrev: 5’ CCGGC*GAATTC*CTGATATCGGGGTA 3’

4-6-2 Sequencing primer

pTAT-HA fwd: 5’ ATC TCG ATC CCG CGA AAT TA 3’

5- Bacteria, bacteriophage and plasmids

M13K07 bacteriophage, used as helper phage was obtained from Pharmacia Biotech, Freiburg, Germany and amplified according to a protocol described below.

Bacteriophage

M13KO7	Km^R-Gen (Tn903) plasmid replication origin (p15A) integrated in replication of origin.	Viera and Messing, 1987

***E. coli* strains**	**Genotype/Phenotype**	**References**
BL21(DE3)pLysS	E.coli, B, F^-, *dcm*, *ompT*, *hsdS*(r_B-m_B-), *gal* λ(DE3) [pLysS, Cam^r]	Studier et al., 1990
JM101λ	supE, thi-1, Δ(*lac- proAB*), [F', *traD36*, *pro*AB, *lacI*^{q}ZΔ*M15*], λ^+	Yanish-Perron et al., 1985
WK6	*gal*E, *str*A, *nal*T, Δ(*lac*, *pro*), F'[*pro*AB, *lacI*^{q}ZΔ*M15*]	Zell and Fritz, 1987
WK6λ	*gal*E, *str*A, *nal*T, Δ(*lac*, *pro*), F'[*pro*AB, *lacI*^{q}ZΔ*M15*], λ^+	Zell and Fritz, 1987 Stanssens et al., 1989
WK6λmutS	*gal*E, *str*A, *nal*T, Δ(*lac*, *pro*), mutS::Tn10, F'[*pro*AB, *lacI*^{q}ZΔ*M15*], λ^+	Zell and Fritz, 1987 Stanssens et al., 1989

Plasmids	**Genotype/Phenotype**	**References**
pSKAN8	*PSTI-pIII*-fusion (λP_L promotor). ori fd; ori ColE1; Amp^R, 5,6 kb.	Röttgen and Collins, 1995
pROCOS4/7	refer to pSKAN8. integration of a cos site, from pTL5 vector. *PSTI* gene removed.	Röttgen and Collins, unpublished
pGEX-5X3	*tac* promoter, *lacI*^{q} gene, Amp^R, factor Xa recognition site, GST as fusion partner, 4,9 kb.	Smith and Jonhson, 1988
pTAT-HA	T7 promotor; Amp^R; His Tag; Hemaglutinin tag; TAT peptide as fusion partner, 3kb.	Nagahara et al., 1998

6- Media and antibiotics for growth of bacteria

6-1 Media

LB-Medium (1x)

Trypton	10 g
Yeast-Extract	10 g
NaCl	5 g
H_2O	to 1 l, autoclaved

LB-Top agar

LB-Medium (1x)	1 l
Agar-Agar	6g, autoclaved

M9 salt solution (10x)

$Na_2HPO_4*2H_2O$	74,1 g
KH_2PO_4	30 g
NaCl	5 g
NH_4Cl	10 g
H_2O	to 1 l, autoclaved

M9-Medium (1x)

M9 salt solution (10x)	100 ml
Glucose(40% w/v)	12,5 ml
CaCl (1M)	100 µl
$MgSO_4$ (1M)	1ml
$FeCl_3$ (1mM)	0,5 ml
Thiamin (10mg/ml)	100 µl
H_2O	to 1 l

All solutions are autoclaved separately or filtered (0.2 µm) for sterilization (Thiamin).

Agar stock solution (2X)

Agar-Agar	15 g
H_2O	to 500 ml; autoclaved

For the preparation of agar plates, 500 ml agar stock solution (2x) were carefully (with the lid partially opened) heated in a microwave until the solution was boiling, mixed with 2X LB-Medium or M9-Medium and allowed to cool down to approximately 50°C. Appropriate antibiotic(s) were added, and after additional mixing, the solution was poured into petri dishes.

6-2 Antibiotic stock solutions

All stock solutions of the antibiotics were stored at -20 °C.

Ampicillin (50 mg/ml)

Ampicillin (sodium salt)	2,5 g
Ethanol (70% in H_2O)	50 ml

Kanamycin (50mg/ml)

Kanamycin	2,5 g
H_2O	50 ml, filter (0.2 µm)

Tetracyclin (20 mg/ml)

Tetracyclin	1 g
Ethanol (70% in H_2O)	50 ml

Selection for a suitable resistance marker was achieved by using the following final concentrations: Ampicillin (Amp) 300 µg/ml

Tetracyclin (Tc) 20µg/ml

Kanamycin (Km) 50 µg/ml

7- Buffers and solutions

Standard stock solutions and buffers that are used for many different purposes are described in this section.

Unique buffers or solutions used only for one particular task are given in the context of the corresponding technique.

7-1 Buffers

PBS (10x)

NaCl	80 g
KCl	2 g
$Na_2HPO_4*2H_2O$	14,3 g
KH2PO4	2 g
H_2O	to 1 l

Adjusted to pH 7 with HCl and autoclaved

TBS (10x)

Tris	61 g
NaCl	80 g
KCl	2 g
H_2O	to 1 l

Adjusted to pH 7,4 with HCl and autoclaved

TE (1x)

Tris-HCl	10 mM
EDTA	1mM

Adjusted to pH 8,0 with HCl and autoclaved

Tris-HCl

Tris	1 M

Adjusted to pH 8,0 with HCl and autoclaved

TA (25x)

Tris	121,1 g
EDTA	18,6 g
H_2O	to 1 l

Adjusted to pH 8,0 with HAc and autoclaved

7-2 Solutions

Phenol

Crystalline phenol was melted, and extracted with Tris-HCl (1 M, pH 8.0) until the water phase had a pH between 7 and 8. A spatula-point of 8-hydroxyquinoline was added to the organic phase. The organic phase was stored protected against light at -20 °C.

Alternatively, Roti Phenol was used (Carl Roth GmbH&Co.KG, Karlsruhe, Germany).

Chloroform-Isoamylalcohol (24/1)

Chloroform	240 ml
Isoamylalcohol	10 ml

EDTA-stock solution

EDTA	0,5 M

Adjusted to pH 7,0 with HCl and autoclaved

Ethidium Bromide (5mg/ml)

Ethidium bromide	50 mg
H_2O	10 ml

Stored protected against light

PEG/NaCl (16,7%/3,3 M)

PEG 8000	100 g
NaCl	116,9 g
H_2O	475 ml

APS (Ammoniumpersulfat solution; 10% w/v)

APS	1 g
H_2O	10 ml

Stored in 500 µl aliquots at -20 °C

Sodium azid stock solution (2,5%)

NaN_3	2,5 g
H_2O	100 ml

Stored at room temperature protected against light

8- Antibodies

8-1 Primary antibodies

Anti-pIII: mouse, monoclonal IgG1	Tesar et al. (1995)
Anti-M13: mouse monoclonal, HRP labeled	Pharmacia Biotech Europe, Freiburg, Germany
Anti-HA: rabbit, polyclonal IgG	Santa Cruz Biotechnology Inc, Santa Cruz, California, USA

8-2 Secondary antibodies

Goat anti-mouse IgG + IgM (H+L), HPR labeled	Dianova GmbH, Hamburg, Germany
Goat anti-rabbit IgG, HPR labeled	Santa Cruz Biotechnology Inc, Santa Cruz, California, USA

B- Methods

1- Work with bacteria

1-1 Growth of bacteria

E. coli strains were grown in 0,5 to 2 liters culture baffled flasks containing LB-medium. Where applicable the suitable antibiotic was added. The incubations were performed at 37°C, under air agitation conditions (180 rpm).

1-2 Storage of bacteria

Individual colonies of all strains were obtained by striking the pertaining strain out on agar plates. Strains carrying an F' factor were spread out on M9 plates (containing the suitable antibiotic). Other strains were grown on LB agar plates, containing the appropriate antibiotic. Incubations were performed at 37°C. The plates were stored at 4°C for short time periods (less than 2 weeks).

For long time storage, 500 µl of a fresh overnight culture were mixed with 500 µl sterile glycerol (87%) in sterile tubes. The suspension was frozen with liquid nitrogen and stored at -70°C.

2- General methods to handle DNA

2-1 DNA agarose electrophoresis

Loading sample buffer

40% saccharose (w/v)

0,25% bromophenol blue

Plasmid DNA, DNA fragments and PCR products were separated on horizontal agarose gels, both in analytical and preparative scale.

TA (1x) was used as a running buffer. The 0,6 - 4,5% agarose gels were prepared by dissolving 0.6 to 3 g of standard agarose (molecular biology grade, Gibco BRL life technologies Ltd, Paisley, United Kingdom) in 100 ml 1x TA buffer by heating the suspension of the agarose in TA. The agarose solution was allowed to cool to approximately 60°C and 5 µl of ethidium bromide solution (10 mg/ml; Sigma-Aldrich Chemie GmbH, Deisenhofen, Germany) were added. Agarose was poured into appropriate gel chambers containing a comb for the formation of sample pockets, was allowed to polymerize before being covered with TA buffer.

Depending of the comb chosen, sample volumes ranging from1 to 20 µl were loaded on the gel after mixing them with 0,5 vol of 2x loading sample buffer. A suitable marker was loaded. Migrations take place at 60 to 120 V, considering the recommendations of Dube (1997). The different bands were detected on an UV transilluminator.

2-2 Phenol extraction of DNA solutions

Contaminant proteins within DNA solutions were removed by phenol extraction. An equal volume of phenol/chloroform/isoamyl (25/24/1) was added to the solution, and the mixture was vortexed for 1 min and centrifuged for 5 min in a desktop centrifuge (13000 rpm). The top aqueous phase was transferred into a fresh tube and the phenol layer was back extracted with an equal volume of TE (vortex and centrifuge). Both top aqueous layers were pooled.

Any trace of phenol was removed by carrying out a chloroform extraction, by adding an equal volume of chloroform/isoamyl (24/1), and then, by mixing and spinning as for a phenol extraction. The top aqueous phase was transferred into a fresh tube. Further purification was achieved by DNA precipitation.

2-3 DNA precipitation

DNA was precipitated by adding 0,2 volume 3 M sodium acetate (pH 5.2), and 2,5 volume of ice cold 100% ethanol. After mixing, the solution was placed at -20°C for 30 min (or 2 hours at -70°C for small DNA). DNA was pelleted by centrifugation at 13000 rpm, at least 30 min (4°C). The supernatant was discarded, and the pellet carefully washed with 70% ice cold ethanol, to remove salts, and centrifuged for further 15 min (13000rpm at 4°C). The supernatant was discarded, and the pellet was air dried at room temperature, and dissolved in appropriate volume of H_2O or TE.

2-4 DNA concentration

DNA concentration was determined with the GeneQuant II RNA/DNA Calculator (Pharmacia Biotech Europe GmbH, Freiburg, Germany) with the 100µl Hellma quartz suprasil glass cuvette (10 mm).

After washing the cuvette thoroughly, a background absorption was determined by measuring the absorption of autoclaved double deionized H_2O or appropriate buffer. DNA to be measured was diluted in the solution used as blank. DNA concentration was obtained by multiplying the absorbance (A260 nm) by the factor of dilution, and by the nucleic acid conversion factor:

DNA Concentration [µg/ml] = A(260 nm) x dilution factor x F

where F is the conversion factor: 50 (ds DNA), 33 (ss DNA), 40 (RNA).

When the purity of the isolated DNA was questioned, absorbance of the DNA solution was measured at 260 and 280 nm. The A(260)/A(280) quotient (DNA/proteins ratio) should range between 1,8 and 2,0 for pure products. If the quotient exceeds 2, the presence of phenol or proteins seemed to be likely and the DNA solution was purified a second time.

3- Methods for *in vitro* recombination of DNA

3-1 Plasmid DNA isolation

Plasmid DNA was isolated with a method based on alkaline lysis, described by Birnboim and Doly (1979).

Small scale preparation (up to 20 µg) of plasmid DNA was performed with the ion exchange columns of Qiagen in the QIAprep Spin Miniprep Kit (Qiagen GmbH, Hilden, Germany). Larger amounts (up to 500 µg) of plasmid DNA were prepared with the Qiagen

Plasmid Maxiprep Kit (Qiagen GmbH, Hilden, Germany) according to the recommendations of the manufacturer.

3-2 DNA gel extraction from agarose gels

For the isolation of a particular DNA from a mixture of DNA fragments, the DNA solution was loaded and fragments were separated on agarose gel through migration. The agarose containing the DNA fragment of interest, was cut out on an UV transilluminator with a scalpel. The DNA extraction was performed with the Qiaquick Gel extraction kit (Qiagen GmbH, Hilden, Germany) according to the manufacturer guidelines.

3-3 Removal of biotinylated DNA

The Dynabeads M-280 streptavidin (Dynal AS, Oslo, Norway) are designed as a matrix for simple and efficient separations of biotinylated compounds (such as DNA, RNA, proteins, immunoglobulins, lectins, sugars). In the course of this work, dynabeads coated with streptavidin, were used to remove the biotinylated restriction fragments of PCR products to obtain pure PCR products that can be further ligated into the appropriate vector.

The dynabeads were washed to remove the preservative NaN_3, by placing the tube containing the beads to be washed, in a Dynal MPC (Magnetic Particle Concentrator; Dynal AS, Oslo, Norway) for at least 30 sec. The supernatant was removed by aspiration with a pipette, and the beads were resuspended in TE/NaCl (2M). The washing step was repeated a second time.

The DNA to be purified (PCR products, amplified with biotinylated primers, and digested on their extremities by restriction enzymes) was mixed with an equal volume of dynabeads M-280 streptavidin, containing a suitable amount of the magnetic beads, and incubated for 30 min at room temperature, by gentle rotation of the tube. 1 mg of dynabeads M-280 streptavidin, binds up to 40 pmol of an approximately 300 bp double stranded DNA fragment.

The purified DNA was recovered by placing the tube containing the mixture in the Dynal MPC for at least 30 sec. The biotinylated fragments, captured by the beads, were attracted by the magnet, and stayed on the wall of the tube which interacts with the Dynal MPC. The supernatant, containing the clean PCR products, was recovered by aspiration with a pipette.

3-4 DNA modifications

3-4-1 Polymerase chain reaction

PCR was used for the amplification of DNA fragments. Oligonucleotides flanked by appropriate endonuclease sites were used as primers. DNA was amplified, based on the protocol of Sambrook et al. (1989), with Taq-DNA-polymerase and DNA-polymerase buffer from Promega (Promega Corp., Madison, WI, USA). The reactions were performed in 0,2 ml PCR reaction tube (Biozym Diagnostik GmbH, Oldendorf, Germany), using DNA Thermal Cycler (Perkin Elmer GmbH, Weiterstadt, Germany). This cycler contains a heated lid and the use of mineral oil was avoided.

PCR reaction:

Oligonucleotide	10 pmol
Sense primer	200 pmol
Antisense primer	200 pmol
DNA-polymerase buffer (10x)	10 µl
$MgCl_2$ (25mM)	6 µl
dNTP (2 mM)	10 µl
Taq-DNA-polymerase	2U
H_2O	to 100 µl

94°C	1 min	10 cycles
60°C	1 min	
72°C	1 min	

The performance of each PCR reaction was checked by running 5 µl of each reaction on agarose gels, with appropriate DNA markers.

3-4-2 DNA digestion

DNA was digested with endonucleases from MBI Fermentas (MBI Fermentas Inc, Amherst, NY, USA) and New England Biolabs (New England Biolabs Inc, Beverly, MA, USA) using the recommended buffers.

For double digestion the optimal buffer was determined experimentally. When appropriate, the double digestion was performed in two following restriction steps, with

modification of the buffer.

The DNA solution was mixed with 1/10 of the final volume of digestion buffer (10x), and with the appropriate amount of endonucleases (1 Unit of endonuclease was used for 1 μg of DNA). The mixture was incubated at 37°C (except for SmaI, which requires 30°C to be active) for at least 2 hours.

For sensitive restriction enzymes, star activity was avoided by diluting the total DNA/restriction enzyme solution in a final volume exceeding 10 times the volume of added endonucleases, and by avoiding long incubation times (no overnight incubation).

When PCR products with cleavage sites in direct vicinity to the termini of the strand, were digested, the amount of endonuclease was increased up to 20 U/μg of DNA and the incubation time at 37°C, prolonged to 4 hours.

Each digestion was verified by electrophoresis. According to the further utilization of the DNA of interest, it was phenol extracted, purified through the GFX PCR DNA and gel band purification kit (Amersham Pharmacia Biotech, Freiburg, Germany), or separated from other components by gel electrophoresis and extracted from the gel (Qiaquick Gel extraction kit; Qiagen GmbH, Hilden, Germany).

3-4-3 Dephosphorylation

To avoid ring closure of digested opened vectors during standard ligation reactions, in addition to the use of two different restriction enzymes, the 5' phosphate groups of the vector were removed by alkaline phosphatase. The digestion mixture was incubated with the appropriate amount of shrimps alkaline phosphatase (1 Unit/pmol of DNA; USB corp., Cleveland, USA), and this mixture was incubated for 1 hour at 37 °C. The shrimps alkaline phosphatase was heat inactivated at 60°C for 15 min.

3-4-4 Ligation

DNA fragments were ligated with T4 DNA ligase (Boehringer Mannheim, Germany) using the provided ligation buffer. Double stranded DNA molecules with complementary cohesive termini that base pair with one another and bring together 3’-OH and 5’-P termini, are ligated by the T4 DNA ligase, through the formation of a phosphodiester bond between adjacent 3’-OH and 5’-P termini.

Vector and insert were ligated together at various molar ratio (vector : insert = 2:1; 1:l; 1:2; 1:3; 1:5), to determine the proportions which lead to the optimal ligation. One Unit of T4 DNA ligase was added to 0,5 µg of DNA to be ligated. This mixture was incubated at least 4 hours at 4°C.

After the ligation, for an increased electroporation efficiency, the ligated DNA was heated up at 65°C for 15 min (Ymer, 1991; Koch Michelsen, 1995).

3-5 Transformation

The transformation of *Escherichia coli* cells was performed through electroporation, according to Dower et al. (1988), using a Bio-Rad Gene Pulser (Bio-Rad laboratories GmbH, Munich, Germany).

3-5-1 Preparation of electrocompetent cells

HEPES solution

1mM HEPES dissolved in H_20 (pH 7)

Autoclaved

10% Glycerol

(87%) glycerol	11,5 ml
H_2O	to 100 ml

Autoclaved

Electrocompetent *E. coli* cells were prepared according to a protocol, adapted from Dower et al. (1988). The strains to be transformed, in the course of this work, were WK6λmutS and BL21(DE3)pLysS.

The strain was spread out on M9 agar plates, with appropriate antibiotics. A single colony was used to inoculate 20 ml of LB (with appropriate antibiotic) and grown at 37°C, at 180 rpm overnight. Two flasks with 500 ml LB (with appropriate antibiotic) were inoculated at 1% with the preculture, and grown at 37°C/180 rpm until the OD(600) ranged between 0,55 and 0,6. The flasks were chilled on ice for 15 min, and cultures were transferred to 4 sterile, ice-cold Sorvall GS3 centrifuge tubes, and spinned in a Sorvall RCSC centrifuge (GS3 rotor) for 15 min at 5000 rpm. After careful discarding of the supernatant, the pellets were resuspended in 4 x 250 ml ice cold HEPES solution, and spinned for further 15 min at 5000

rpm. After decanting the supernatant the pellets were resuspended in 4 x 125 ml ice cold H_20, and pooled in two suspensions. This was followed by a further centrifugation for 15 min at 5000 rpm (G53, RCSC), the supernatant was decanted and the *E. coli* cells resuspended in 2x 10 ml 10 % glycerol. After transfer into SS34 centrifugation tubes the cells were pelleted by a 10 min centrifugation (6000 rpm, SS34 rotor), the supernatant was discarded and the cells resuspended in 2x 1 ml of 10 % glycerol. The suspension was aliquoted, frozen in liquid nitrogen and stored at -70°C. All steps were carried out as quick as possible at 4°C.

Each batch of electrocompetent cells was checked by electroporating 10 ng of pBR322. Dilution series of the control were plated on LB plates with appropriate antibiotics.

3-5-2 Electroporation

The electrocompetent cells were carefully mixed (by flicking) with 1 to 2 µl of a ligation mixture (approximately 10 ng of DNA) and incubated on ice for 2 min. This mixture was transferred into ice cold 2 mm electroporation cuvette (Bio-Rad laboratories GmbH, Munich, Germany) and were electroporated with the following parameters: 2,5 kV, 25 µF and 200 Ω, as described by Dower et al. (1988). The time constant was controlled, and ranged between 4,5 and 4,7 depending on the purity of the DNA (salt).

1 ml of prewarmed LB (37°C) was added and mixed to the cells **immediately** after the electroporation, using a sterile pasteur pipette, and transferred into a sterile 1,5 ml Eppendorf tube. This suspension was grown at 37°C for one hour.

In order to evaluate the electroporation efficiency, and therefore, the number of transformants obtained, dilution series were prepared from an aliquot of the suspension, plated out on LB plates containing appropriate antibiotics and incubated at 37°C overnight.

For the generation of phagemid libraries, several electroporations were set up in parallel, to obtained a large diversity library. Transformed bacteria resulting from each electroporation mix, were plated out on two large LB plates.

3-6 DNA sequencing

Two sequencing protocols were applied, using the ALF DNA sequencer (Pharmacia Biotech Europe GmbH, Freiburg, Germany), or the Applied Biosystem ABI prism 310 sequencer (Perkin Elmer Applied Biosystems GmbH, Weiterstadt, Germany). Both systems work with fluorescence labeling. However, with the ALF DNA sequencer, DNA fragments are separated through migration along a gel, while for the ABI prism 310, separation occurs

within a capillary, fill in with polymer.

3-6-1 ALF-DNA-Sequencer

3-6-1-1 Cycle sequencing reaction

Cycle sequencing buffer (10X)

40 mM $MgCl_2$

150 mM $(NH_4)_2SO_4$

120 mM Tris-HCl (pH 9.5)

dNTP-Mix

dATP, dCTP, dTTP and 7deaza-dGTP: 1 mM each

ddNTP-Mix

ddATP, ddCTP, ddTTP: 5µM each

ddGTP: 3,75 µM

Stop solution .

95% formamide

10 mM NaOH

0.005 % bromophenolblue

The protocol used here, is adapted to the sequencing of double stranded DNA.

For each DNA probe to be sequenced, four cycle sequencing reactions are performed in 0,2 ml PCR reaction tubes. In each of the four tubes, 2 µl of one of the ddNTP was added (ddATP, ddCTP, ddGTP, or ddTTP). The following 20 µl mix was then added to each of the tube:

Cycle sequencing buffer (10X)	2,0 µl
fluorescein labeled primer (2 pmol/µl)	1,0 µl
dNTP-Mix (1 mM, each)	3,0 µl
Taquenase/PPase (Amersham, 32 U/µl)	0,16 µl
plasmid DNA	0,5 to 1 µg
ddH_2O	to 20 µl

The PCR-reaction was run as followed:

	94°C	2 min	
	94°C	15 sec	35 cycles
	46°C	15 sec	
	72°C	40 sec	
	72°C	5 min	

After completion of the PCR program, the reaction was stopped by adding 4 µl of stop solution to each reaction. Samples were stored at -20°C, or directly sequenced.

3-6-1-2 Preparation of Sequencing Gels

TBE (10x)

Tris	100 g
H_3BO_3	55,6 g
EDTA	9,3 g

ALF DNA-sequencer gel plates were cleaned by washing with SDS (10 %) solution, ethanol and purified water (millipore). The gel cassette was assembled according to the ALF-manual.

6,6 % gel of 0.5 mm thickness was prepared by mixing:

urea (ALF grade)	33,6 g
TBE (10X)	9,6 ml
acrylamide/bis (40%;29:1)	13,2 ml
ddH_2O	to 80 ml

The mix was degased by vacuum filtration, TEMED (65 µl) and APS (10%; 240 µl) were added, an the solution was immediately poured between the glass plates and the gel comb was inserted. The gel was allowed to polymerize for at least 2 hr. The comb was removed and the slots were cleaned. The gel cassette was assembled within the ALF DNA sequencer, according to the ALF-manual, and TBE (0,6x) was poured within the tanks on both extremities of the gel.

Immediately before loading, the DNA samples were heated up to 95°C for 3 min, and chilled on ice. 8 µl of each DNA samples were loaded onto the gel. The gel was run according to the ALF manual, and the results were analyzed automatically by the ALF DNA sequencer software.

3-6-2 Applied Biosystems ABI Prism 310

On the ABI 310 the dye terminator thermo cycle sequencing method was used to sequence DNA fragments. Plasmid DNA to be sequenced were isolated and purified with the QIAprep Spin Miniprep Kit (Qiagen GmbH, Hilden, Germany).

In 0.2 ml PCR strip tubes 200 to 500 ng of plasmid were mixed with 3 pmol of the appropriate sequencing primer, 4 µl of big dye terminator premix (Big Dye Terminator Ready Reaction Cycle Sequencing Kit with AmpliTaq FS; PE Applied Biosystems, Weiterstadt, Germany), and this mixture was filled up to 20 µl with water. The resulting mixtures were subjected to PCR sequencing reaction.

96°C	30 sec	25 cycles
50°C	15 sec	
60°C	4 min	

After completion of the PCR sequencing reaction, the mixture was purified by ethanol precipitation. All steps were performed at room temperature. The 20 µl reactions, were mixed with 2 µl of NaAc (3M; pH 5,2) and 50 µl 100% ethanol (at room temperature), and mixed by inverting several times the tubes. Samples were incubated 10 min, and centrifuged 30 min on a table centrifuge, at maximum speed. Pellets were washed with 250 µl of 70% ethanol (room temperature) and a further 20 min centrifugation was performed. Pellets were allowed to air dry for 10 min, and were resuspended in 12 µl TSR reagent (PE Applied Biosystems, Weiterstadt, Germany). The solution was heated up to 95 °C for 2 min and chilled on ice. The DNA was stored at -20°C, or directly sequenced.

The capillary of the ABI prism 310 sequencer was filled up with POP6 polymer, and sequencing could take place.

4- Protein analysis

4-1 Determination of protein concentration

The concentration of the proteins was determined according to Bradford (1976). A calibration curve was performed with BSA. 160 µl of fixed concentration of BSA were mixed in a microtiter well, with 40 µl of Bio-Rad reagent (Bio-Rad laboratories GmbH, Munich, Germany). A negative control with 160 ml of water and 40 ml of Bio-Rad reagent was done, to evaluate the background value. After 5 min incubation, the OD_{595} was determined, and the calibration curve was drawn (concentration versus OD_{595}), after deduction of the background value from the protein values. The calibration curve should be linear in the range of 20 to 140

µg.

A dilution series of the protein to be analyzed was done (in water), and 160 µl of each dilution was mixed in a microtiter well, with 40 µl of Bio-Rad reagent. After 5 min, the OD_{595} was determined, and the concentration of the protein was evaluated by reporting the OD value to the calibration curve. To obtain a precise concentration, the evaluation was achieved from the average of triplicates, taken into account the dilution factor.

4-2 SDS-Polyacrylamide gel electrophoresis (SDS-PAGE)

Upper Tris

Trisbase	30,5 g
ddH_20	to 400 ml; adjust with concentrated HCl to pH 6,7
+ SDS (10%)	20 ml
+ ddH_2O	to 500 ml

Lower Tris

Trisbase	91 g
ddH_20	to 400 ml; adjust with concentrated HCl to pH 8,8
+ SDS (10%)	20 ml
+ ddH_2O	to 500 ml

Running buffer (8x)

Trisbase	560 g
Glycin	120 g
SDS (10%)	200 ml
EDTA (0,5 M)	80 ml
ddH_2O	to 2500 ml

Sample buffer (2x)

Tris (100 mM) pH 6,8
DTT (200 mM)
SDS (4%)
Bromphenol blue (0,2%)
Glycerin (20%)

Separation gels, for Biometra 8,5 x 7,5 x 0,1 cm gel (for 15 µl sample volume)

	15%	**12,5%**	**10%**
acrylamide/bis (30%)	5 ml	4,1 ml	3,3 ml
Lower Tris	2,5 ml	2,5 ml	2,5 ml
ddH_20	2,4 ml	3,2 ml	4,1 ml
Temed	8 µl	8 µl	8 µl
APS (10%)	50 µl	50 µl	50µl

Stacking gel, for Biometra 8,5 x 7,5 x 0,1 cm gel (for 15 µl sample volume)

acrylamide/bis (30%)	700 µl
Upper Tris	2 ml
ddH_20	4,2 ml
Temed	7 µl
APS (10%)	35 µl

SDS-Polyacrylamide gels (Laemmli, 1970) were used to evaluate the quality, and purity of proteins.

The suitable separation gel was poured between fixed cleaned glass plates (10% SDS and ethanol) to approximately 1 cm under the comb and covered with a layer of water. After 20 min of polymerisation the water was removed.

The stacking gel was poured on top of the separating gel. A suitable comb was inserted and the gel allowed to polymerize (30 min). The gel was placed in the electrophoresis unit, and covered with the running buffer. The comb was removed, and the sample pockets were washed with running buffer.

1 to 7 µl of the samples were mixed with equal volumes of the 2x sample buffer, heated up to 95°C for 3 min, and chilled on ice, together with the molecular weight standards. Samples were loaded onto the gel, which was run at 60 V to 100 V until appropriate separation occurred.

Gels were further used for protein detection by coomassie blue or silver staining or by western blotting antibodies detection.

4-3 Staining methods

4-3-1 Coomassie blue staining

Coomassie staining solution

Coomassie brilliant blue G250: 1 g in 200 ml ethanol

Coomassie brilliant blue R250: 1 g in 200 ml H_2O

+ Glacial acidic acid 40 ml

Filtration

Proteins in PAGE gels were detected after electrophoresis, by incubation for 30 to 60 min in a coomassie staining solution, at room temperature under constant rocking. Surplus stain was removed from the gel by boiling it in water, on a microwave, until the protein bands got clearly visible and the background minimized.

4-3-2 Silver staining

Solution A

Methanol	100 ml
Acid acetic	20 ml
ddH_2O	100 ml

Solution B

Methanol (50%)	100 ml
TCA (12%)	24 g
$CaCl_2$ (2%)	4g
ddH_2O	to 200 ml

Solution C

Isopropanol (10%)	20 ml
Acid acetic (5%)	10 ml
ddH_2O	170 ml

Solution D

$KMnO_4$ (0,01%)	0,02 g
ddH_2O	200 ml

Solution E

$AgNO_3$ (0,1%)	0,2 g
ddH_2O	200 ml

Solution F

Formaldehyde (37%)	500 µl
Na_2CO_3 (2%)	10 g
ddH_2O	to 500 ml

Silver staining procedure (Heukeshoven and Dernick, 1988) is more sensitive than the coomassie blue technique. It is therefore suitable for the detection of low protein amounts.

The gel was incubated on a rocker platform with the freshly prepared following solutions:

1- twice in solution A, 15 min each time.

2- 15 min in solution B

3- 10 min in solution C

4- 5 min in solution D

5- 10 min in solution C

6- Twice for 10 min in H_2O

7- 10 min in solution E

8- 10 sec in H_2O

9- the gel was developed in 100 ml of solution F for 2 min. The solution was discarded, and the gel was once more in fresh solution F incubated, for a further 2 min, and so on, until the bands are clearly visible.

10- the reaction was stopped with 10 min incubation in 1% acid acetic, and then, rinse in water.

4-4 Western blot

4-4-1 Blotting of polyacrylamide gels on nitrocellulose

Semidry blotting buffer

Tris	48 mM
Glycine	39 mM
SDS	0,037%
Methanol	20%

T-TBS buffer

Tween 20 (0,05%)	0,5 ml
TBS	995,5 ml

The transfer of proteins from SDS-PAGE gel nitrocellulose membrane was performed through the use of a semi dry blotting apparatus (PHASE, Lübeck, Germany).

Four sheets of suitable sized Whatman 3 MM paper (Whatman, Maidstone, England) were soaked in semidry blotting buffer and placed on the blotting chamber. One dry sheet of Whatman paper was used to pull off the polyacrylamide gel from the glass plate, was moistened, and placed on top of the other sheets (gel above). A suitable sized nitrocellulose membrane (HAHY 0.45 mm pores, Millipore, Eschborn, Germany), was moistened in semidry blotting buffer, and placed above the gel, on the blotting chamber. Air bubbles were avoided. Five moistened sheets of Whatman paper were placed above the nitrocellulose.

Potential air bubbles were removed by rolling a 10 ml glass pipette over the Whatman papers. The lid of the blotting chamber was assembled with the bottom part and a currency of 300 mA was applied for 2 hours.

After completion of the transfer, the chamber was disassembled, and the nitrocellulose was subjected to an immuno-assay.

4-4-2 Immunological detection

T-TBS (1x)

TBS (10x)	100 ml
Tween 20	500 µl (0,05%)
ddH_2O	to 1000 ml

Blocking solution

T-TBS (1x)	100 ml
Skimmed milk powder	2 g (2%)

Staining solution (freshly prepared)

T-TBS	20 ml
Diaminobenzidine (2%)	200 µl
NiCl2 (10%)	200 µl
H_2O_2 (30%)	20 µl

After the transfer of the proteins to the nitrocellulose, the membrane was blocked for 30 min at room temperature, with 2% dried skimmed milk powder (Magermilchpulver, Gluckskle, Germany) in T-TBS, to cover the part of the nitrocellulose that are free of protein. The blocking solution was discarded, and the membrane was washed once with T-TBS buffer.

The proteins were incubated at least 2 hours with the first antibody, diluted to the suitable concentration in blocking solution. The membrane was washed twice for 5 min with T-TBS, and incubated for 2 hours with an appropriate horseradish peroxidase labeled secondary antibody, diluted in T-TBS. The membrane was intensively washed with T-TBS.

Detection of the bound antibody, the membrane was incubated with freshly prepared staining solution for 5 to 30 min at room temperature, until the intensity of the bands was appropriate. The membrane was then rinsed in water to stop the reaction.

5- Preparation and characterization of bacteriophage

5-1 Preparation of M13KO7 helper phage

5-1-1 Production of M13KO7 plaques

E. coli WK6 was spread out on a M9 plate and grown at 37°C. A single colony was picked to inoculate an overnight culture from which an 20 ml LB (in a 100 ml baffled flask) culture at 1% was started. The culture was incubated at 37°C and 180 rpm until the OD_{600} reached 0,5.

A 500 µl aliquot of the WK6 culture was infected with 1 µl of M13K07 bacteriophage (around $9x10^{11}$ pfu/ml, Pharmacia Biotech Europe, Freiburg, Germany), and incubated 15 min at 37°C, to allow the phage particles to infect the cells. 3 ml of LB Top agar (pre-melted, and kept warm at 47°C to prevent solidification) was added to the cell suspension, gently mixed together, and immediately poured onto a plate containing hardened bottom agar medium. Swirl the plate gently, to ensure a even distribution of bacteria and top agar. The top agar was allowed to harden at room temperature for30 min, and the plate was inverted and incubated overnight at 37°C.

5-1-2 Preparation of a M13KO7 stock

A well separated plaque was picked using a sterile pasteur pipette equipped with a rubber bulb. The pipette stab through the chosen plaque into the hard agar beneath. A mild suction was applied so that the plaque, together with the underlying agar, was drawn into the

pipette. The plaque was added to 20 ml of LB (2x) containing 50 µg/ml of kanamycin, and was grown for 6 to 8 h at 37°C/180 rpm. This overday culture was used to inoculate 4 flasks containing 250 ml of LB (2x)/Kanamycin (50µg/ml) at 1%, which was allowed to grow overnight at 37°C/180 rpm.

The bacteria were then pelleted by centrifugation (15 min at 8000 rpm, 4°C, GS3 rotor). The supernatant was transferred to a fresh bottle, and centrifuged a second time, under the same conditions. The supernatant was then mixed with 0,15 volume of PEG/NaCl, and incubated on ice for at least 2 hours. The phage were precipitated by 90 minute centrifugation at 4°C, 8000 rpm (GS3 rotor). The supernatant was decanted and completely removed after a second brief centrifugation (3 min at 5000 rpm, GS3 rotor). The pellet was dissolved in 10 ml of PBS (1x), cleared by 10 min centrifugation at 12000 rpm and 4°C (SS34 rotor). NaN_3 was added to the resuspension, at a final concentration of 0,02%, and the M13KO7 stock was stored at 4°C. The phage titer usually ranged between 10^{11} and 10^{12} cfu/ml.

5-2 Packaging of phagemid DNA libraries in bacteriophage particles

After insertion of the hypervariable domains within the pSKAN8 or pROCOS4/7 vectors, the libraries were created by transforming a large number of WK6λmutS bacteria with the phagemid vectors.

After electroporation, transformed *E. coli* WK6λmutS were plated out on large LB agar plate, containing 300 µg/ml ampicillin, and 20 µg/ml tetracycline, and allowed to grow overnight at 37°C. The colonies were then scraped of all agar plates, by adding 15 ml LB to each plate. Cells were allowed to resuspend by gently rocking the plates for 20 min at room temperature, before washing them by pipetting the 15 ml LB up and down with a 10 ml pipette. The cells were transferred into a sterile falcon tube, and the plates were rinsed with an additional 5 ml of LB. Recovered cells were mixed by inverting the tube several times. 2 ml of this suspension were used to inoculate 200 ml LB medium containing 100 µg/ml of ampicillin. The remaining suspension was mixed with 0,5 volume of sterile glycerol, frozen in liquid nitrogen, and stored at -70°C.

The 200 ml culture was allowed to grow until OD_{600} reached 0,5.

Around 10^{11} cfu of M13KO7 (stock solution) were added to the logarithmic phase WK6λmutS, and the culture was incubated at 37°C without shaking for 30 min, to allow the regeneration of the F-pili, and to let time to the helper phage to infect the cells. The culture was then grown overnight at 37°C/180 rpm.

The cells were pelleted and discarded by centrifugation (15 min, 8000 rpm, 4°C, GSA rotor). The supernatant was transferred into a fresh bottle and centrifuged a second time, under the same conditions. The phage suspension was then mixed with 0,15 volume of PEG/NaCl, and incubated on ice for at least 2 hours. The phage were precipitated by centrifugation (1 hour, 8000 rpm, 4°C, GSA rotor). Trace of medium were eliminated by a brief centrifugation (3 min at 5000 rpm), and the phage pellet was dissolved in 1 ml PBS. The suspension was cleared by centrifugation (10 min on a bench centrifuge, 13000 rpm, room temperature), and NaN_3 were added to a final concentration of 0,02%. The titer has to be determined.

5-3 Determination of bacteriophage titers

20 ml of LB medium containing 20µg/ml of tetracycline were inoculated with 200µl of WK6λmutS overnight culture, and incubated at 37°C and 180 rpm until OD_{600} reached 0,5.

A dilution series of the phage to be titered was performed: 90 µl of sterile water was placed in the wells of a standard sterile 96 well microtiter plate and the dilution series was prepared by transferring 10 µl aliquot of the phage preparation (dilution 0 to 10^{-10}).

100 µl of the logarithmic growing WK6λmutS bacteria (OD_{600}=0,5), were added to the dilution series, carefully mixed, and incubated for 30 min at 37°C.

20 µl of each dilution were spotted on a LB agar plate containing ampicillin (300 µg/ml) and tetracycline (20 µg/ml) to evaluate the titration of the phagemid, and on a LB agar plate containing kanamycin (50 µg/ml) and tetracycline (20 µg/ml) to estimate the helper phage titer. The colonies in each spots were counted, and allowed the determination of the cfu titer (colony forming units per ml), taking into account the dilution factors.

6- Affinity selection from the phagemid libraries

6-1 Panning procedure

T(0,05%)-PBS (1x)

PBS (10x)	100 ml
Tween 20	500 µl
ddH_2O	to 1000 ml

Elution buffer

Glycine (0,1M; pH 2,2)

Blocking solution

T(0,05%)-PBS (1x)	100 ml
Skimmed milk powder	2 g (2%)

Neutralization buffer

Tris (2M)

A standard protocol has been designed to isolate and enrich for our particular SH3 and EVH1 domain targets, a few particles of a whole library, which present feature(s) required for binding to the target. The target domains were immobilized and incubated with the library. Phagemids displaying peptides containing a motif essential for interaction with the target were retained during washing steps, and then eluted and used for reinfection of WK6λmutS, to be amplified. After superinfection with helper phage, the phagemid particles were subjected to subsequent panning cycles.

Affinity selections were performed in microtiter plates which were subjected to a special treatment of the polystyrol surface, in order to increase the affinity for polar groups. The chosen plates were the maxisorb microtiter plates (Nunc A/S, Roskilde, Denmark), except when specified (Reacti-Bind Glutathione coated plates; Pierce Chemical Company; Rockford; Illinois, USA). 1 µg of the GST fusion proteins of the SH3 and EVH1 domains used as target proteins and resuspended in PBS buffer (100 µl), were immobilized overnight at 4°C, on constant gently rocking, on the wells of the microtiter plate. The wells were previously sealed with parafilm strips (American National Can™, Chicago, IL, USA). This sealing is repeated for each incubation step.

The supernatant was removed and the plate washed thoroughly with T(0,05%)-PBS. The wells of the plate were blocked with 400 µl of the blocking solution, for 30 min at room temperature, on constant gently rocking, to prevent unspecific binding of the phage to the walls of the wells, and to keep background levels low. After washing with T(0,05%)-PBS, the phagemid libraries diluted 1:1 in blocking solution (phage which show affinity for milk, would interact with milk in solution and not with immobilized milk, and therefore, would be, later, washed away more easier), were incubated with the immobilized target for at least 2 hours at room temperature, with constant rocking (usually between 10^{10} and 10^{11} phage/well).

The supernatant was removed, and the wells were washed with several cycles of T(0,05%)-PBS, and alternatively, with blocking solution. The stringency of the washing steps increased with the cycles of the panning selection. Refer to Table II-1 to see the washing conditions used in affinity selection (when other parameters were used, the conditions were specified directly in the section of interest). The wells were finally rinsed once with water, to remove trace of Tween 20.

	cycle of panning		
	1	2	3 - 5
T(0,05%)-PBS	1x	3x	5x
PBS/milk	10 min	10 min	10 min
T(0,05%)-PBS	1x	3x	5x
PBS/milk	5 min	5 min	5 min
T(0,05%)-PBS	1x	3x	5x
H_2O	1x	1x	1x

Table II-1: Stringency of the washing steps according to the cycle of panning. Basic procedure for the washing step. The PBS/Tween contains 0,05% of Tween 20. The PBS/milk, is the blocking solution, with 2% of skim milk powder. Incubation with the blocking solution tends to reduce the background level. It can happen that phage bind to the milk powder used to block the walls of the wells, the milk added in solution competes this interaction.

The remained bound phage were eluted with 100 µl of elution buffer, after an incubation of 15 min at room temperature on constant rocking. The solution was then neutralized by transferring the eluted phage into a Eppendorf reaction tube containing 6 µl of Tris (2M). A 10 µl aliquot of eluted phage were used to evaluate the titer of the recovered phage, while the remaining 90 µl were used to infect the *E.coli* strain.

6-2 Reinfection of WK6λmutS with the eluted phage

A culture of WK6λmutS was incubated at 37°C, until the OD_{600} reached 0,5 (the culture volume depends on the number of targets used during the panning; 10 ml of culture were required for eluted phage from each target).

The 90 µl of eluted phage were mixed with 10 ml of the log-phase WK6λmutS and were incubated for 30 min at 37°C. The bacteria were then pelleted by centrifugation (8000 rpm; 5 min; SS34 rotor) and resuspended in 400 µl LB medium containing ampicillin (300 µg/ml)

and tetracycline (20 µg/ml). Each suspension was plated out on LB/Agar (ampicillin: 300 µg/ml; Tetracycline: 20 µg/ml) plates, and grown overnight at 37°C.

6-3 Packaging of phagemids from reinfected cells

The bacteria lawn was resuspended by adding 15 ml of LB medium to each plate and by gently rocking the plates for 20 min at room temperature, before washing them by pipetting the 15 ml LB up and down with a 10 ml pipette.

1 ml of the resuspended cells, was used to inoculate 25 ml LB (ampicillin: 300 µg/ml; tetracycline: 20 µg/ml). The resuspension was incubated for 1 hour at 37°C and 180 rpm. Superinfection was performed by adding 10^{11} to 10^{12} cfu of M13KO7 helper phage, and incubating the mixture without shaking for 15 min at 37°C, and then, for an additional 5 hours at 37°C, and 180 rpm

The cells were pelleted and discarded by centrifugation (10 min, 5000 rpm, 4°C, SS34 rotor). The supernatant was transferred into a fresh bottle and centrifuged a second time, under the same conditions. The phage suspension was then mixed with 0,15 volume of PEG/NaCl, and incubated on ice for at least 1 hour. The phage were precipitated by centrifugation (15 min, 18000 rpm, 4°C, SS34 rotor). The phage pellet was dissolved in 200 µl of PBS. The suspension was cleared by centrifugation (10 min on a bench centrifuge, 13000 rpm, room temperature), and NaN_3 were added to a final concentration of 0,02%.

The phage were ready for a further cycle of panning selection. After a few rounds of panning selection, enrichment of eluted phage might be detected through the titration of the given and eluted phage after each cycle. The enriched clones were characterized by sequencing randomly picked single clones.

6-4 Characterization of selected single clones

6-4-1 Single clone packaging

A 3 ml LB (ampicillin: 300 µg/ml; tetracycline: 20 µg/ml) overnight culture were inoculated with a single WK6λmutS colony containing the phagemid of interest.

This preculture was used to inoculate a 100 ml LB (ampicillin, tetracycline) at 1%, which was incubated at 37°C and 180 rpm, until OD_{600} reached 0,5. Superinfection was performed by adding around 10^{11} M13KO7 helper phage, and incubating the mixture for 15 min at 37°C without shaking, followed by overnight incubation at 37°C and 180 rpm.

The single phagemid particles were then recovered as previously described: the cells

were eliminated by two following centrifugation steps (10 min, 8000 rpm, 4°C, GSA rotor). The supernatant was transferred in fresh bottle, mixed with 0,15 volume of PEG/NaCl, and incubated at least two hours on ice. Phage were pelleted by centrifugation (90 min, 10000rpm, 4°C, GSA rotor), complete trace of medium were removed, and the phage pellet was resuspended in 1 ml of PBS. The suspension was cleared by centrifugation (10 min, 13000 rpm, room temperature, table centrifuge). NaN_3 was added to a final concentration of 0,02%. The titer of each propagated single clone was determined, and the suspension was stored at 4°C.

6-4-2 Phage ELISA

T(0,5%)-PBS (1x)

PBS (10x)	100 ml
Tween 20	5 ml
ddH_2O	to 1000 ml

Blocking solution

T(0,5%)-PBS (1x)	100 ml
Skimmed milk powder	2 g (2%)

Staining solution

Na_2HPO_4 (200 mM)	3,1 ml
Citric acid (200 mM)	2,9 ml
ddH_2O	6 ml
O-phenylenediamine	4,8 mg
H_2O_2	1,5 µl

To check the interaction between the enriched clones (determined by sequencing) and the target, the clones were propagated as single phagemid particles, and tested on phage ELISA (Enzyme Linked Immuno Sorbent Assay).

1 µg of the proteins of interest were diluted in 50 to 100 µl PBS, and coated on a microtiter plate (Nunc-MaxiSorb plates; Nunc A/S, Roskilde, Denmark) by incubation overnight at 4°C (or alternatively, 2 hours at room temperature).

The supernatant was slapped out of the wells, which were then wash once with T(0,5%)-PBS, before being fill in with the blocking solution and incubated for 1 hr at room temperature. The wells were then washed once with T(0,5%)-PBS, and the phage preparation diluted 1:1 in blocking solution, were incubated with the immobilized proteins for at least 1,5 hour (10^{10} phage per well).

The wells were washed twice with T(0,5%)-PBS and incubated for 1 hour with 100 µl of αM13-antibody (horseradish peroxydase labeled; Pharmacia Biotech Europe GmbH, Freiburg, Germany) diluted 1:2000 in T(0,5%)-PBS.

The immobilized proteins were intensively washed with T(0,5%)-PBS, and 100 µl of the staining solution were added to the wells. After 5 to 30 min, when the coloration was appropriate, the reaction was stopped by adding 50 µl H_2SO_4 (2 M), and the absorbance at 492 nm was measured, using an ELISA-plate reader.

6-4-3 Competitive phage ELISA

Competitive phage ELISA were performed according to the same principle than the phage ELISA. But on this case, the interaction between the immobilized protein, and the phage displayed peptide, was studied in presence of increasing amount of competitor. In the course of this work, the used competitor is S1, the highest affinity peptide displayed on phage surface, for Src SH3 domain, isolated by cosmix plexing library, and synthesized as free peptide.

The procedure was similar to the one of phage ELISA, the phage preparation were allowed to bind to the immobilized protein, but after washing, the free peptide competitor was added to the phage particles attached to the immobilized protein. Alternatively, the phage preparation, and the free peptide competitor were incubated simultaneously with the immobilized protein.

Like for the phage ELISA, the amount of remained bound phage was then evaluated by using anti-M13 antibody (horseradish peroxydase labeled; Pharmacia Biotech Europe GmbH, Freiburg, Germany).

7- Determination of constant affinity of the selected variants

The BIAcore 2000 (Pharmacia Biosensor, Uppsala, Sweden) was used to test isolated peptides for binding to their SH3 domain targets, and for determination of affinity constant.

7-1 Immobilization chemistry

Each ligand to be immobilized was diluted in buffers of pH 0,5 to 3 pH units smaller than the isoelectric point of the ligand of interest. pI were determined with the ExPASy Internet service (http://expasy.hcuge.ch/ch2d/pi_tool.htlm).

Two sensor chips, F1 and SA (Pharmacia Biosensor, Uppsala, Sweden), were used to immobilized the target proteins, using conventional amino coupling.

The F1 sensor chip was used to immobilized the GST-SH3 domain. A continuous flow of PBS/Tween 20 was maintained over the sensor surface. The carboxyl groups on the sensor surfaces were activated with an injection of a solution containing 0,2 M N-ethyl-N'-(3-diethylamino-propyl) carbodiimide (EDC) and 0,05 M N-hydroxysuccinimide (NHS). The F1 dextran matrix was covered with antibody against GST. This was reached by injecting the antibody. The immobilization procedure was completed by a 7 min injection of 1 M ethanolamine hydrochloride, to block remaining ester groups. Capture of the GST-SH3 domains by the antibody against GST was obtained by injection of GST-SH3 domain diluted in PBS/Tween. This yielded an amount of 2000 to 5000 RU of bound protein. This method was relevant to characterize the ligands able to interact strongly to their target.

The SA sensor chip, coated with streptavidin, was used to immobilized short biotinylated peptides, according to the EDC/NHS chemistry. 10 to 30 RU of biotinylated peptides were bound to the chip by carefully injecting small volumes of the peptides by using the manual inject command. This strategy is convenient to determine the interaction kinetic constants.

For all immobilization protocols, one reference channel (flow cell 1) was not coated with the target protein, but treated like the three other flow cells. This reference cell was used for background correction.

7-2 Binding of the interaction partners

GST-SH3 domains were immobilized on the F1 sensor chip, trapped with the anti GST antibody. The phage preparations were diluted in PBS/Tween 20, and injected as continuous flow over the sensor surface, at a flow rate of 5 µl/min. The time for association was 6 min and the dissociation was studied for 8 min. The sensor chip was regenerated by two washings with 1 M glycine pH 2,2, for 30 sec, at a flow rate of 20 µl/min. A following cycle was then

performed, where GST-SH3 domains were immobilized on the anti GST antibody covering the F1 sensor chip, and other concentration of the phage preparation were injected to the chip.

As far as the SA sensor chip was concerned, the GST-SH3 domains were added in the flow through buffer, to interact with the 10 to 30 RU of bound biotinylated peptides. Concentrations of GST-SH3 domain which ranged from 0,3 to 10 nM were injected. Conditions of injection were identical to the one of F1 sensor chip, except for the regeneration, with was performed by washing with 10 mM NaOH, for 30 sec, at a flow rate of 20 µl/min. The affinity constants were calculated for each interaction, as described under the section III-G of results.

8- Preparation of the Tat fusion peptides

After insertion of the peptide of interest within the polylinker site (*Xho*I and *Eco*RI recognition sites) of the pTAT-HA vector (kindly provided by Steven F. Dowdy; University of Washington, Saint Louis, Missouri, USA), the recombinant vectors were transformed in BL21(DE3)LysS strain (Stratagen, La Jolla, CA, USA).

8-1 Purification of the Tat-fusion peptides

Urea	8 M

Buffer Z

Urea	8 M
NaCl	100 mM
Hepes	20 mM (pH 8,0)

Imidazole (500 mM)

Imidazole	17,02 g
Buffer Z	to 500 ml

A 500 ml LB culture of the transformed cells was inoculated, and incubated for 4 hours at 37°C, pelleted, washed once in PBS, and finally resuspended in 8 M urea, to denature the fusion proteins.

After sonication, purification of the Tat fusion protein was performed, taking advantage of the His tag, by applying the suspension to a Ni-NTA column (Qiagen GmbH, Hilden, Germany), previously equilibrated with buffer Z. The suspension applied to the Ni-NTA column was first diluted in buffer Z/10 mM Imidazole. The column was then washed with buffer Z completed with 10 mM Imidazole, to minimize non specific binding, and elution was done by applying 500 mM Imidazole diluted in buffer Z, to the column. The fractions were analyzed on Coomassie blue stained SDS-PAGE gel, and western blot, where Tat fusion peptides were detected with anti hemaglutinin antibody.

8-2 Ionic exchange chromatography

An ionic exchange chromatography was performed, to eliminate the urea from the protein suspension.

Buffer A

NaCl	50 mM
Hepes	20 mM (pH 8,0)

Buffer B

NaCl	1 M
Hepes	20 mM (pH 8,0)

NaCl (1 M)

NaCl	11,7 g
ddH_2O	to 200 ml

The suspension was applied to a HiTrap SP 1 ml column (Pharmacia Biotech Europe GmbH, Freiburg, Germany) previously equilibrated with 5 ml buffer A, followed by 5 ml of buffer B, and once more with 5 ml buffer A. All buffers were applied with a flow through of 1 ml/min.

The column was washed with 5 ml of buffer A, and the proteins were eluted with 3 ml of 1 M NaCl.

The Tat-fusion peptides were then desalted on a PD 10 disposable G-25 sephadex gravity column (Pharmacia Biotech Europe GmbH, Freiburg, Germany). The PD 10 column was equilibrated with 25 ml PBS, and the protein suspension was loaded. After the volume

entered the gel bed, more PBS was applied, and fractions were collected, and analyzed on Coomassie blue stained SDS-PAGE gel.

8-3 Fluorescein labeling

The Tat fusion peptides were conjugated with fluorescein isothiocyante (FITC; λ excitation: 494; λ emission 520); Pierce, Rockford, IL, USA), to facilitate the detection of the Tat-peptides delivered to the cells.

FITC stock solution (daily prepared)

FITC	1 mg
DMSO	500 µl

Kept in dark

10x FITC conjugation buffer (prepared monthly)

$NaHCO_3$ 1 M/ Na_2CO_3 1 M

Adjust the pH to 9,0 by adding solution B to solution a until pH reaches 9,5.

Solution A

$NaHCO_3$	0,84 g
ddH_2O	to 10 ml

Solution B

Na_2CO_3	1,06 g
ddH_2O	to 10 ml

The Tat peptides to be conjugated with FITC (usually between 5 and 25 µg) was added to 30 µl of 10x conjugation buffer, 10 to 50µl of FITC stock solution, and ddH_2O was added to reach a final volume of 300 µl. The reaction mixture was incubated for two hours at room temperature, in the dark.

The FITC-labeled Tat fusion peptides were loaded to a PD 10 disposable G-25 sephadex gravity column (Pharmacia Biotech Europe GmbH, Freiburg, Germany), as described above, section II-8-2.

The fluorescein labeled proteins were stored at –70°C with 10% glycerol.

9- Transduction of cells with Tat fusion peptides

9-1 Cell culture

The cells used for the Tat-fusion peptide transduction were the NIH 3T3 fibroblasts (Andersson et al., 1979; Copeland et al., 1979). They were cultured at 37°C, on a humidified air of 10% CO_2.

Trypsin/EDTA

Trypsin (2,5%)	20 ml
EDTA (500 mM)	0,6 ml
PBS	480 ml

FCS: inactivated for 30 min at 56°C

Pen/Strep (100x)

Penicillin	10000 U/ml
Streptomycin	10 mg/ml

Completed DMEM (Dulbecco's Modified Eagle's Medium)

DMEM	445 ml
Pen/Strep (100x)	5 ml
Inactivated FCS	50 ml

NIH-3T3 fibroblasts grow as an adherent monolayer of cells.

The medium of the cultured cells was changed each time it was required, in average every 2 to 4 days. Cultures were split at 1:10 every 4 to 6 days.

Passage of adherent cells was performed when the cells reached a confluent stage. The medium was carefully discarded, the cells were washed once with PBS, and recovered with a layer of Trypsin/EDTA solution. After a few minutes, the cells were detached from the surface. The trypsin was then inactivated. The neutralization was performed by adding at least the double amount of medium. The cells were then pelleted by centrifugation, and resuspended in fresh medium to a suitable dilution, prior to be placed in fresh flasks.

For long term storage, NIH-3T3 cells were resuspended in 70% DMEM, 20% FCS and 10% DMSO ($1x10^6$ cells/ampoule). The suspension was frozen with liquid nitrogen and stored at -70°C.

9-2 Trypan blue exclusion test of cell viability

This technique is based on the principle that viable cells have intact cell membranes and exclude dyes such as trypan blue whereas dead cells do not. Thus, viable cells have a clear cytoplasm while non-viable cells are stained in blue.

Equal volumes of the cell suspension and of 0,4% Trypan blue solution (Sigma-Aldrich Chemie GmbH, Deisenhofen, Germany) were mixed. After 1-2 min incubation, 10 µl were transferred into the chamber of a Neubauer haemocytometer and the cells counted using a binocular microscope.

Each corner square of the haemocytometer has 16 small squares and has 1 mm sides; since the volume of liquid covering each corner is 0,1 mm, the total volume in the field is 10^{-4} cm^3. Since 1 cm^3 is equivalent to 1 ml, the subsequent cell concentration per ml was determined using the following calculation:

No. of cells per ml = average count per square x dilution factor x 10^4

9-3 Transduction of NIH 3T3 cells with Tat-fusion peptides

NIH 3T3 cells to be transfected with Tat fusion peptides were trypsinized from their cultured flasks, and resuspended at a concentration of 10^6 cells/ml of fresh medium completed with FCS.

1 ml cell suspension was placed on a non treated surface, where adherence is prevented. FITC labeled Tat-fusion peptides were directly added to the medium, at concentrations ranging between 10 and 160 nM (to check for dose dependent transduction). Several aliquots were prepared, so that there were enough samples to check by Fluorescence Activated Cell Sorting (FACS) analysis for efficient transduction at particular time points (0, 15, 30, 45, 60 min).

Aliquots of transduced cells, after one hour incubation with the FITC-Tat-peptides, were pelleted and resuspended in fresh FCS/medium. They were placed into usual flasks, and directly checked on confocal microscopy for transduction. The flasks were then incubated under normal conditions, and regularly checked on confocal microscopy for a few days.

9-4 Flow cytometry analysis

In Fluorescence Activated Cell Sorting assays (FACS), cells can be sorted according to their size, the integrity of their membrane, or due to their conjugation with fluorochrom.

Cells were transduced with FITC-Tat-fusion peptides, as described above (section II-9-3). After incubation time of interest between the cells and the Tat-peptides, the cells were recovered, pelleted by centrifugation, and resuspended in 200 µl of cold PBS. This washing step allowed the reduction of the background level. The samples were then incubated on ice, and analyzed with FACS.

III - RESULTS

The aim of this work was to isolate strong ligands for SH3 and EVH1 domains. In order to characterize ligands against particular SH3 or EVH1 domains, one needs to express and purify the target proteins under reliable conditions, to obtain pure proteins, free of contaminants. The GST-fusion system is highly advantageous for achieving a very high degree of purity.

A - GST-fusion proteins

Several vectors have been constructed that simplify the purification of foreign polypeptides expressed in *E.coli* (reviewed by Marston, 1986). The pGEX expression vector family (Pharmacia) directs the synthesis of foreign polypeptides, as fused to the GST protein, which can be purified under non-denaturing conditions by glutathione-affinity chromatography (Smith and Johnson, 1988). The SH3 and EVH1 domains studied in this work were all expressed as GST-fusion proteins.

The nephrocystin SH3 domain was the only protein that was produced entirely as GST-fusion protein, in our laboratory, starting from the cDNA (provided by Dr. Hildebrandt, University of Freiburg, Germany). The cDNA was inserted into the pGEX-5X3 within the restriction sites *Sma*I and *Sal*I, on the C-terminal end of the GST (Figure III-1). All other SH3-fusions proteins were provided by Lawrence Quilliam and Brian Kay (University of Wisconsin-Madison, Wisconsin, USA).

WK6λmutS cultures transformed with recombinant pGEX plasmids were submitted to IPTG induction. This step was also performed for all SH3 domains: the nephrocystin previously cloned into the vector, and several other SH3 domains, supplied by Quilliam and Kay (University of Wisconsin-Madison, Wisconsin, USA) as GST-fusion constructs (within pGEX-2T) and ready for the protein expression. After 5 to 7 hours incubation, the cells were harvested and lysed on ice by mild sonication. The supernatant was then ready to be applied to a glutathione sepharose column (Sigma). The fusion proteins were eluted by competition with free reduced glutathione. The Figure III-2A shows the elution profile of the purified GST-fusion protein. All proteins recovered after this treatment are very pure, as illustrated in Figure III-2B, from GST-Src fusion protein, as a typical example.

The EVH1 domains were kindly supplied by Jürgen Wehland (GBF, Braunschweig, Germany) as purified GST-fusion proteins.

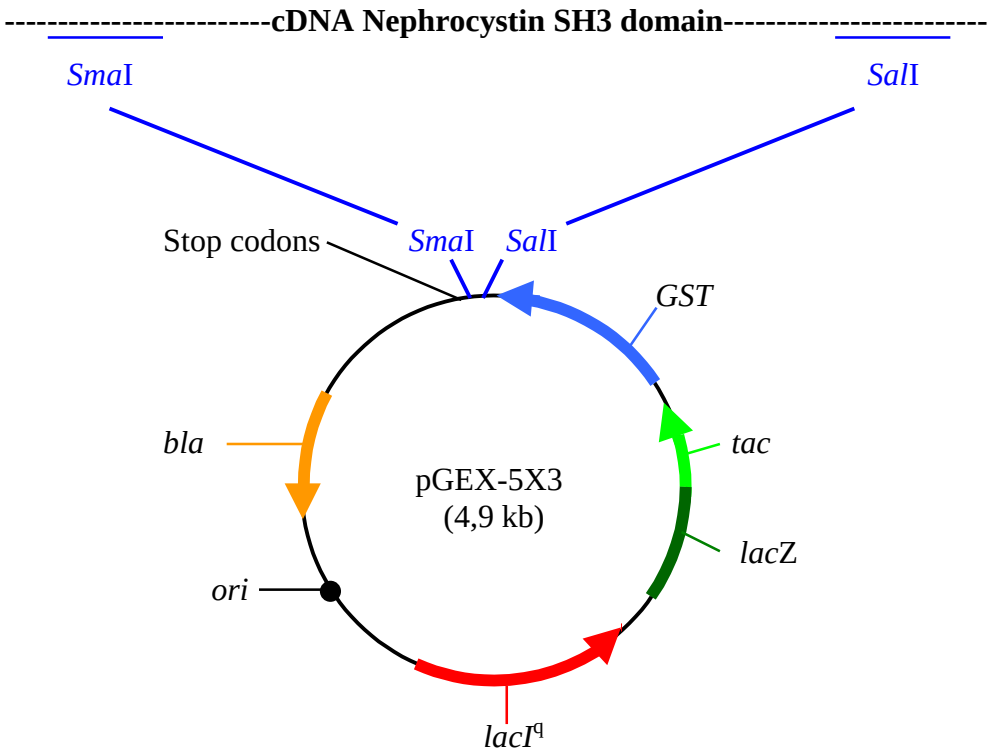


Figure III-1: Schematic map of the plasmid vector pGEX-5X3. The normal translation stop codon of the GST cDNA began at nt 7 and has been destroyed through the introduction of oligodeoxynucleotides encoding unique cleavage sites for the multiple cloning sites and TGA stop codons in all three reading frame. *Sma*I and *Sal*I restriction sites were used to insert the nephrocystin cDNA, C-terminal from the GST. pGEX-5X3 contains additional sequences encoding protease cleavage sites recognized by factor Xa, between the GST and the multiple cloning sites. The GST-fusion gene is under the control of the *tac* promoter. Cells transformed with this plasmid and induced with IPTG, synthesized the GST-fusion protein. pGEX-5X3 carries a fragment of the *lac* operon containing the over-expressed *lacI*^{q} allele of the *lac* repressor and part of *lac*Z. In absence of inducer, the plasmid-encoded *lacI*^{q} allele is efficient in repressing transcription from the *tac* promoter. The following abbreviations are used: *bla*: β-lactamase gene (ampicillin resistance); *ori*: origin of replication.

Both families of proteins (SH3 and EVH1) show preferences for proline rich ligands. Therefore, the libraries generated for the screening were biased libraries (Figure III-3) which contain fixed amino acid residues at particular positions (one or several prolines in each library and, in addition, in two of the libraries, an arginine). By incorporating a structural motif that favors the target binding to a large collection of peptides, one tends to increase the chance of isolating ligands with higher affinity for the target of interest, as well as defined the orientation of the ligands in an uniform way.

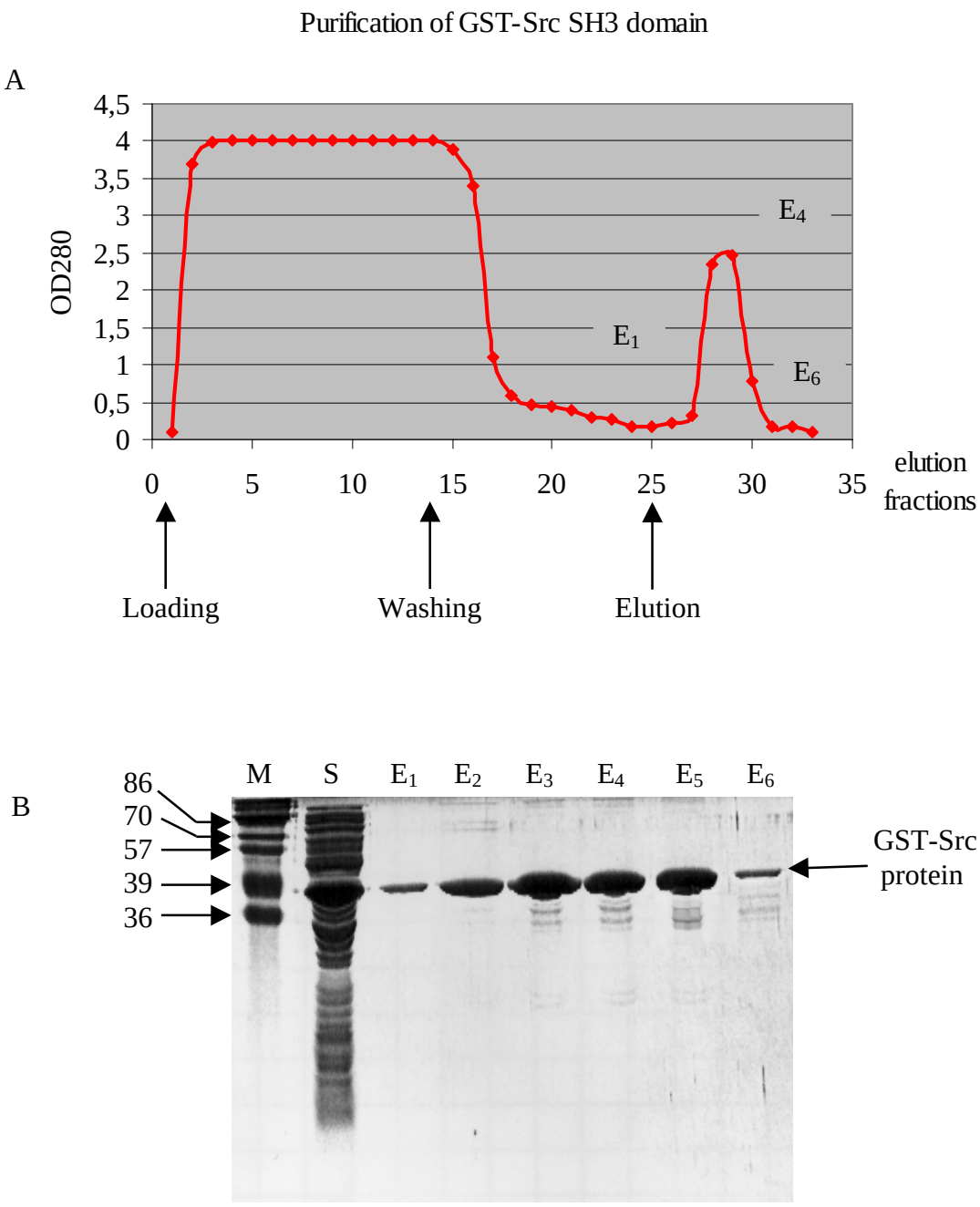


Figure III-2: Purification of GST-Src SH3 domain, on a glutathione sepharose column. (A-) The GST-Src protein loaded onto the column was produced from a 400 ml culture incubated with IPTG (0,1 mM) for 5 hours. Elution fractions (0,5 ml) were recovered throughout the purification procedure, and OD_{280} was used, to determine the protein concentration of each fraction. Washing of the column was performed with equilibration buffer (Tris/EDTA; 10mM/10mM) and was carried out until the eluant reached baseline OD_{280} levels. The GST-Src protein was then eluted by competition with free reduced glutathione. (B-) Elution fractions of the purification of GST-Src domain analyzed by SDS-PAGE, followed by staining with Coomassie blue. S represents the supernatant of the cell extract, as applied to the column. E stands for the eluted fractions, which contain GST-Src protein. The position and sizes (kDa) of molecular weight markers (M) are indicated.

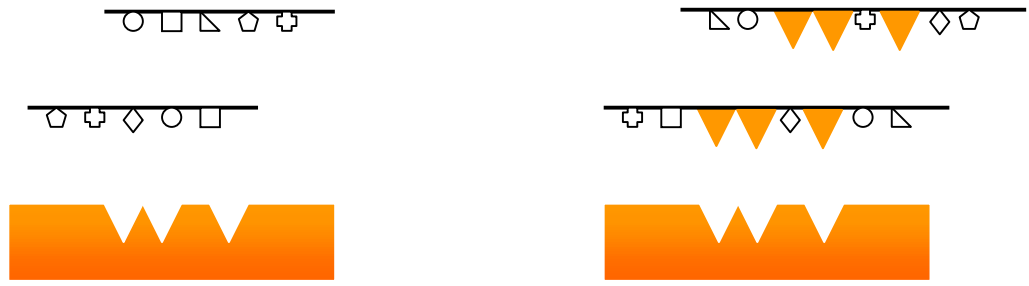


Figure III-3: Comparison of preferences for a given target between a random peptide library and a biased library that contains predetermined ligand-binding preferences. The biasing element tends to increase the average affinity of the ligands for the receptors.

Five different libraries were constructed in two ways. Four banks were constrained, expressing the hypervariable sequence bracketed by two cysteines and therefore, displayed as a loop. The fifth bank, a cosmix-plexing® bank, displayed a linear sequence on the surface of the phage.

The constrained libraries did not result in isolation of ligands for our targets, for reasons that will be discussed below.

B – Constrained random peptide libraries

The constrained libraries were constructed by insertion of synthetic hypervariable oligonucleotides within the vector pSKAN8.

pSKAN8 (Figure III-4) is a phagemid-display vector which contains a pancreatic secretory trypsin inhibitor (PSTI)- pIII fusion gene (Röttgen and Collins, 1995). The PSTI is a 56 amino acid long protein which has been extensively studied in our group: the specificity of the inhibitor was altered to give strong inhibitors (K_i = < 10^{-11} M) for chemotrypsin or human leukocyte elastase (Collins et al., 1990; Szardenings et al., 1990). In the banks derived from the pSKAN8 vector, the hypervariable region replaced a short PSTI loop (amino acids 17 to 23) at the exposed tip of the PSTI. Serine-protease inhibitors of the ovomucoid superfamily, to which PSTI belongs, exhibit rapid evolution mainly via diversity in the constrained loop which is mutated in our banks. This loop is able to accommodate many structures and still fold correctly.

The “outer membrane”-protein, OmpA-leader peptide (Omp_L), was fused to the first amino acid of PSTI to ensure that production, secretion and correct cleavage of the leader

peptide occur (Szardenings et al., 1990).

The PSTI-pIII gene is under the control of the bacteriophage lambda p_L promoter. The λ p_L promoter is strongly repressed by the λ cI protein which is produced in λ lysogen *E. coli* strains. The use of this promoter ensures that the correct ratio of hybrid to wild type pIII in the particles produced on superinfection, is achieved for essentially monovalent presentation (i.e., one hybrid pIII per particle).

The pSKAN8 vector also contains plasmid replication functions, ampicillin and chloramphenicol resistance genes and the M13 phage replication/packaging origin.

The hypervariable domains replacing the 7 amino acid loop (residues 17 to 23) within PSTI employed in creating the four gene banks pSKAN8-HypD, -HypE, -HypF and -HypG were as following:

HypD	Cys	$(Xxx)_4$	**Pro**	Xxx	**Arg**	Cys		
HydE	Cys	**Arg**	$(Xxx)_3$	**Pro**	$(Xxx)_2$	Cys		
HypF	Cys	$(Xxx)_2$	**Pro**	Xxx	Xxx	**Pro**	$(Xxx)_2$	Cys
HypG	Cys	$(Xxx)_3$	**Pro**	**Pro**	Xxx	**Pro**	$(Xxx)_2$	Cys

Each of these banks presents particular advantages in isolating ligands for the SH3 domains, due to the presence of the fixed prolines. In addition, banks pSKAN8-HypF and -HypG might be highly favorable, to characterize EVH1 domain ligands.

The HypD and HydE libraries were especially designed to isolate strong ligands for SH3 domains. Both libraries take advantage of the well known preferences of SH3 ligands of requiring an arginine in order to build a salt bridge with a conserved residue of the SH3 domain (Asp99 in Src).

Two ligand classes for SH3 domains have been reported which bind in opposite orientations to the domain. These two ligand classes share distinct consensus sequences, the main difference between them is the position of the arginine which occurs before or after the proline rich core motif. The class I consensus sequence is RXLPXXP, while it is PXXPXR for the class II. Interestingly, some SH3 domains, like Src or PI3K, can bind peptides of either class (with the consensus sequences RPLPPLP and PPVPPR for class I and II Src ligands, and RXLPPRP and PPLPXR for class I and class II PI3K ligands). Although the biological implications of this are unclear, a possible explanation is that ligand binding in two orientations could access different partners in multi-component protein complexes. HypD and HypE libraries have characteristics (position of the arginine) belonging to class II or class I ligands, respectively.

Primer A

HypD AC*GAGCTC*AACGGT**TGC(NNK)$_2$NNK NNK CCA NNKAGA TGC***GGTACC*GA

Primer B

HypE AC*GAGCTC*AACGGT**TGCAGA(NNK)$_2$NNK CCC NNK NNK TGC***GGTACC*GA

HypF AC*GAGCTC*AACGGT**TGC(NNK)$_2$CCGNNKNNKCCN(NNK)$_2$TGC***GGTACC*GA

HypG AC*GAGCTC*AACGGT**TGC(NNK)$_3$CCACCCNNKCCG(NNK)$_2$ TGC***GGTACC*GA

*Sac*I *Kpn*I

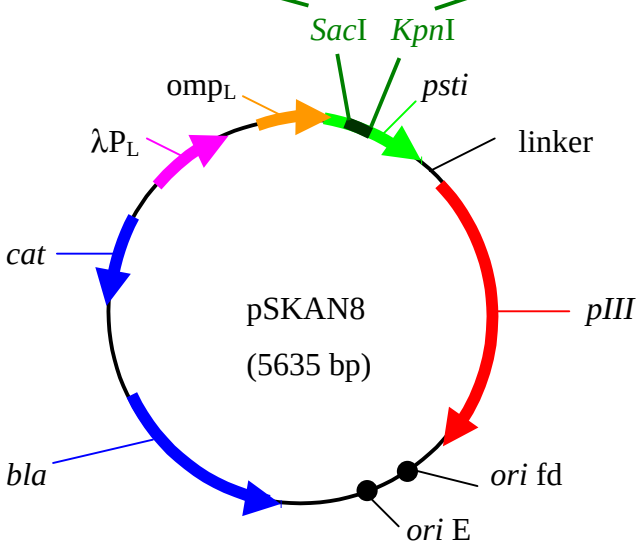


Figure III-4: Schematic map of the phagemid vector pSKAN8. pSKAN8 was described by Röttgen and Collins (1995) for presenting the PSTI on the surface of bacteriophage. The endonuclease sites *Sac*I and *Kpn*I were used to introduce the hypervariable gene banks: HypD, HypE, HypF, HypG. The hypervariable inserts were amplified with the primers A and B, and digested with the restriction enzymes *Sac*I and *Kpn*I. The following abbreviations are used: $\lambda P_{L:}$ major leftward promoter of the bacteriophage λ; omp_L: signal sequence derived from the E.coli outer membrane protein ompA; *psti*: gene encoding for the pancreatic secretory trypsin inhibitor; *pIII*: gene coding for the minor coat protein pIII of the M13 bacteriophage; *bla*: β-lactamase gene (ampicillin resistance); *cat*: chloramphenicol acetyltransferase gene; *ori* fd: origin of replication for single stranded DNA of the bacteriophage fd; *ori* E: origin of replication for plasmids colE1.

Even if Src and PI3K SH3 ligands are known to bind in two different orientations, it is unclear for the other SH3 domain ligands. The work done by B. Kay's group predicts that Yes, Abl and Grb2NH2 SH3 domain ligands tend to bind in a class I orientation, whereas the

peptides selected by the Cortactin, p53bp2, PLCγ, and Crk NH2 SH3 domains are expected to interact in a class II orientation (Sparks, 1996). Therefore, each of the generated banks, HypD and HypE, might be relevant for different SH3 domains.

The HypF and HypG banks are more general and suited to all SH3 and EVH1 domains. They take advantage of the proline rich motif, necessary for the formation of the left-handed polyproline type II helix that the SH3 or EVH1 domain ligands might form in order to bind to these domains.

HypF presents a PXXP core motif which is the essential basic motif of the polyproline type II helix. Each ligand capable of binding to SH3 or EVH1 domains might contain this motif. The PPXP motif displayed in HypG might be selective for SH3 and EVH1 domains, as the presence of a proline will reduce the entropy cost needed for the formation of the polyproline type II helix. Most of the SH3 and EVH1 ligands show this PPXP motif. Hence, of the four constrained banks, HypG may be the one which offers the greatest selectivity. However, HypF offers a larger freedom which may be interesting by giving a larger amino acid selection to the ligand sequences.

1- Generation of the libraries

1-1 Preparation of the linearized vector DNA

The phagemid pSKAN8 stuffer used in this work is an expression vector where variant peptides can be fused to the N terminal part of the pIII gene. Phagemid vectors contain both plasmid and bacteriophage replication origins. However, with the exception of the partner gene for the fusion of the peptide or protein to be displayed, they do not carry any of the phage genes. The phagemids maintain themselves as plasmids, until a superinfection with a helper phage (M13KO7) activates their phage replication/packaging origins. The helper phage provides all bacteriophage functions necessary for the synthesis of the circular single-stranded DNA and for the assembly of the phage-like particles, displaying the hybrid fusion proteins.

The stuffer is a 3 kb DNA fragment, resulting from the digestion of λ DNA with *Ngo*MI and *Hind*III, inserted in the region encoding the PSTI inhibitory loop. On complete digestion with the restriction enzymes (*Sac*I and *Kpn*I), the stuffer and the opened vector are distinct after separation, on an agarose gel. Since two different restriction enzymes are used for the digestion, the gel-extracted large vector fragment has two different cohesive ends. This

fragment cannot undergo ring closure in the absence of an inserted fragment.

The insertion site for the hypervariable oligonucleotide was prepared by linearizing the vector with the *Kpn*I and *Sac*I restriction enzymes. The hypervariable region was inserted between a leader "signal" sequence and the M13*pIII* gene in the phage display vector, so that the peptide would be presented on the surface of the phagemid particle as an amino terminal extension.

1-2 Preparation of the double stranded DNA insert from degenerated oligonucleotides

The codon scheme used to encode random amino acid sequences is NNK, where N is A,C,G or T, and K is G or T. This scheme minimizes the stop codon frequency (which would lead to unproductive clones), as only one of the three stop codons (TAG) can be encoded, while codons for all the 20 amino acids are still represented.

Basic protocols used in many laboratories to clone inserts into a vector lead to the formation of unwanted products, which might seriously limit the electroporation efficiency. The advantage of the method employed here, is its ease of manipulation: vector and insert are digested at both extremities by two different restriction enzymes to prevent ring closure of vector without insert (unproductive product).

Vector and insert are then simply ligated together. However, this technique offers no possibility to direct the formation of the desired product (Figure III-5). At best, one can reduced the proportion of unwanted products by performing a series of test ligations, to determine the ratio of insert to vector DNA which results in the largest number of recombinants after electroporation.

To minimize these difficulties, an optimized strategy for insertion of the hypervariable DNA domain into the vector, was developed. This strategy allows control of the generated products along with the possibility of eliminating small cleavage products, and leads to the formation of a clean product.

Although the strategy chosen to prepare the oligonucleotides involves many steps, however, it guarantees pure products. Figure III-6 illustrates this strategy. The quality of the digested inserts have a strong influence on the efficiency of ligation with the vector and thus, on the efficiency of electroporation of the ligated DNA into bacteria.

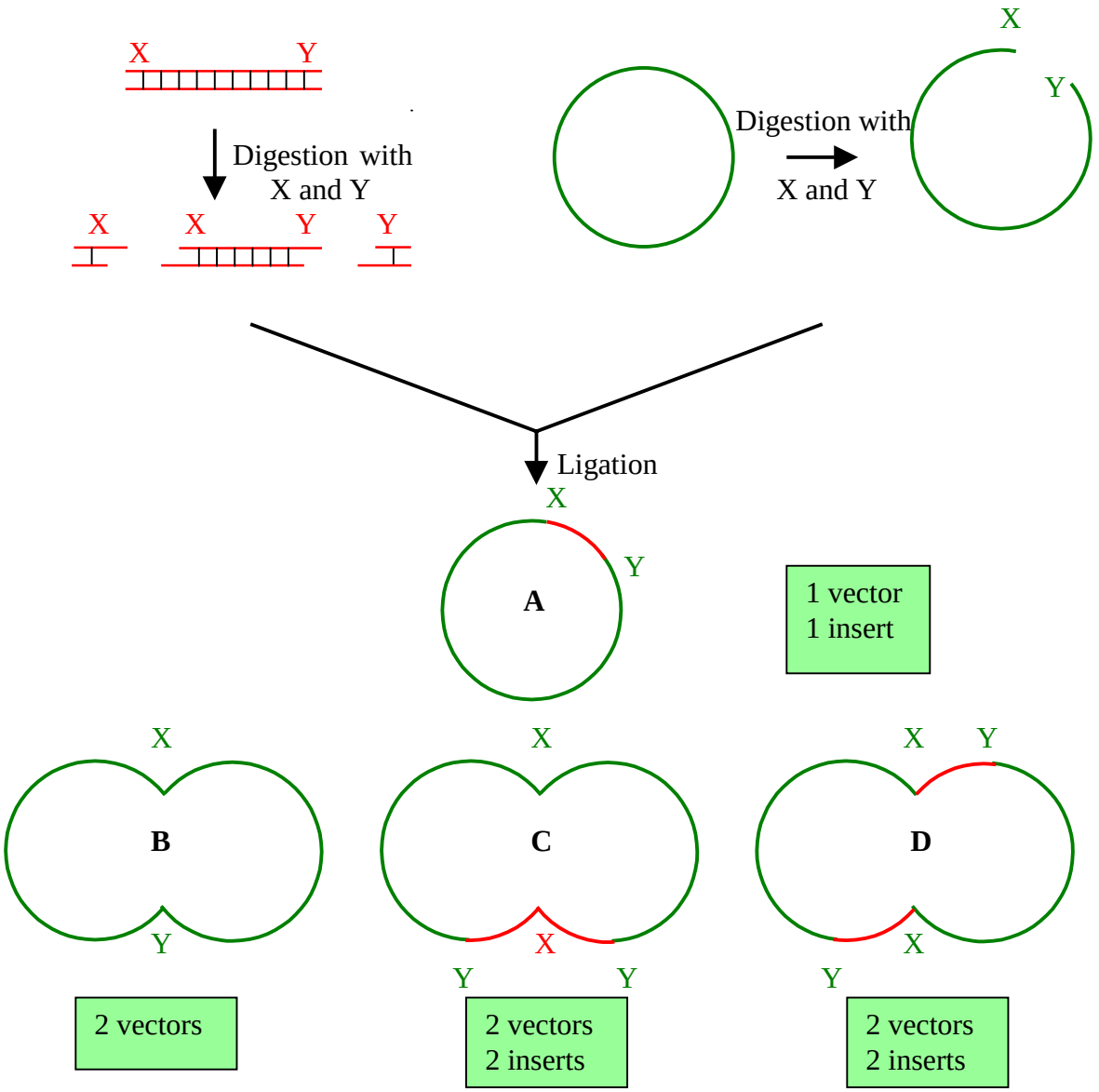


Figure III-5: Basic protocol for clone insertion into vector DNA. Vectors and inserts are digested by two distinct restriction enzymes and ligated together. The ligation step leads to the formation of several products. The vectors which might ligate with small cleavage fragments are not represented. Product A is the expected one: one insert is cloned into the vector. The products B, C, and D leads to unproductive clones. Varying the insert/vector ratio might favor the formation of particular products. The ratio which directs the generation of the largest number of product A, is to be used in a large scale ligation. However, the formation of the other products cannot be completely prevented. If the amount of insert is reduced (for a fixed amount of vector), product B will be favored compared to C and D. Conversely, if an excess of insert is used, products C and D (or vectors which would ligate more inserts) will dominate over B. X and Y, stand for two different restriction enzyme recognition sites.

To minimize these difficulties, we developed an optimized strategy for insertion of the hypervariable DNA domain into the vector. This strategy allows control of the generated products along with the possibility of eliminating small cleavage products, and leads to the

formation of a clean product.

Although the strategy chosen to prepare the oligonucleotides involves many steps, however, it guarantees pure products. Figure III-6 illustrates this strategy. The quality of the digested inserts have a strong influence on the efficiency of ligation with the vector and thus, on the efficiency of electroporation of the ligated DNA into bacteria.

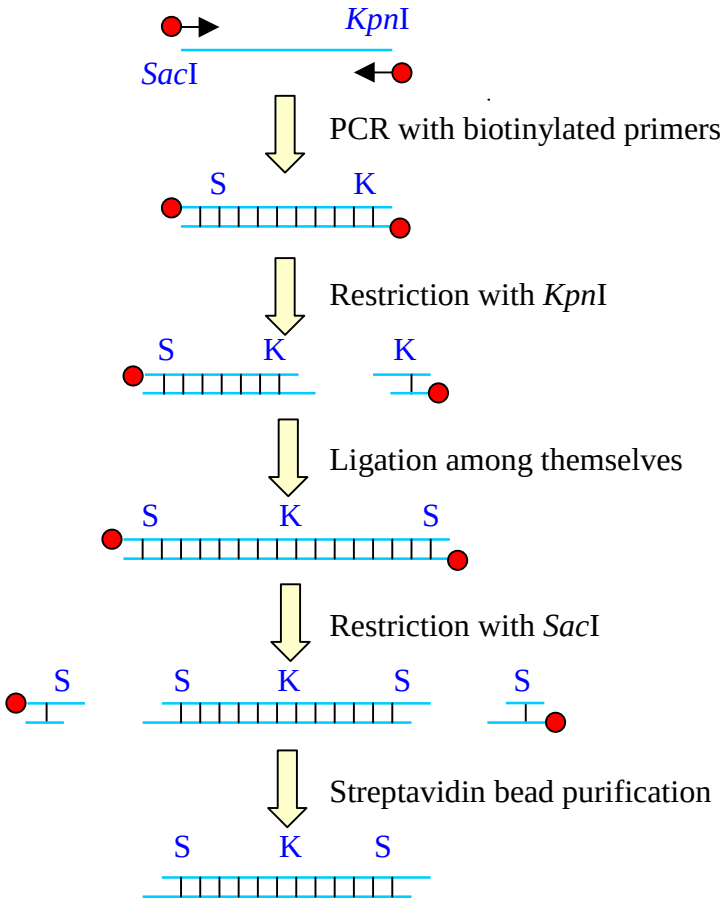


Figure III-6: Schematic representation of insert preparation. The insert oligonucleotides were amplified by PCR using biotinylated primers, leading to products carrying biotin at the extremities. The PCR products were digested with *Kpn*I and ligated among themselves. The resulting dimeric insert was digested with *Sac*I and the restriction fragments were removed by magnetic streptavidin beads. S and K stand for *Sac*I and *Kpn*I sites, respectively.

The insert oligonucleotides were converted into double-stranded DNA and amplified via PCR, using biotinylated primers (to allow subsequent easy removal of restriction fragments). The inserts were then restricted with *Kpn*I, allowed to ligate among themselves to form dimeric inserts with a central *Kpn*I site and, at both extremities, *Sac*I restriction sites. This

fragment was then digested with *Sac*I, generating cohesive ends at both extremities. The product was purified using superparamagnetic polysterene beads with streptavidin covalently attached to the bead surface (Dynabeads® M-280 Streptavidin; Dynal). These beads captured and removed the biotinylated DNA from the digested insert. The insert was ready for ligation with the linearized vector.

1-3 Ligation of the vector and insert DNA

The digested vector and insert were ligated together (Figure III-7). Re-anealment of the vector (formation of parental vectors) was impossible since the ends were different and furthermore, the insert could ligate only with the *Sac*I-digested end of the vector. Restriction with *Kpn*I was performed. The vector attached to one end of the insert was then ligated, leading to the ring closure of the DNA. This ligation was performed at low DNA concentration to encourage ring closure and discourage formation of concatemers or potential reinsertion of the small cleavage products. This protocol led to the formation of a clean product, since it was possible at each step, to direct the formation of the wished particles. The obtained product, reached very high efficiency during electroporation.

The hybrid pSKAN8 thus formed, pSKAN8-HypD, -HypE, -HypF, -HypG, were transformed into WK6λmutS *E.coli* strain (ampicillin selection; 300µg/ml). This lambda lysogen strain expresses the cI transcription repressor of the lambda bacteriophage, and prevents an inopportune production of PSTI-pIII gene product which might be toxic. In addition, it ensures the presentation of one hybrid pIII per particle.

The electroporations were carried out, until relevant bank diversities were reached. For a library containing 5 random residues (pSKAN8-HypD, and -HypE), a diversity equivalent to approximately 10^8 single clones (Table III-1) is necessary to ensure the presence of all possible amino acid combinations, with a 99% level of confidence. For pSKAN8-HypF and -HypG which contain 6 random amino acids, the necessary diversity is about $5x10^9$ single clones. The values of necessary diversities for biological libraries, according to the number of randomized positions are summarized in the Table III-1, adapted from Clackson and Wells observations (1994).

The diversities obtained for pSKAN8-HypD, -HypE, -HypF, -HypG libraries, are presented in Table III-2, and show that the diversities of the libraries are sufficient to represent most of the possible random peptide combinations.

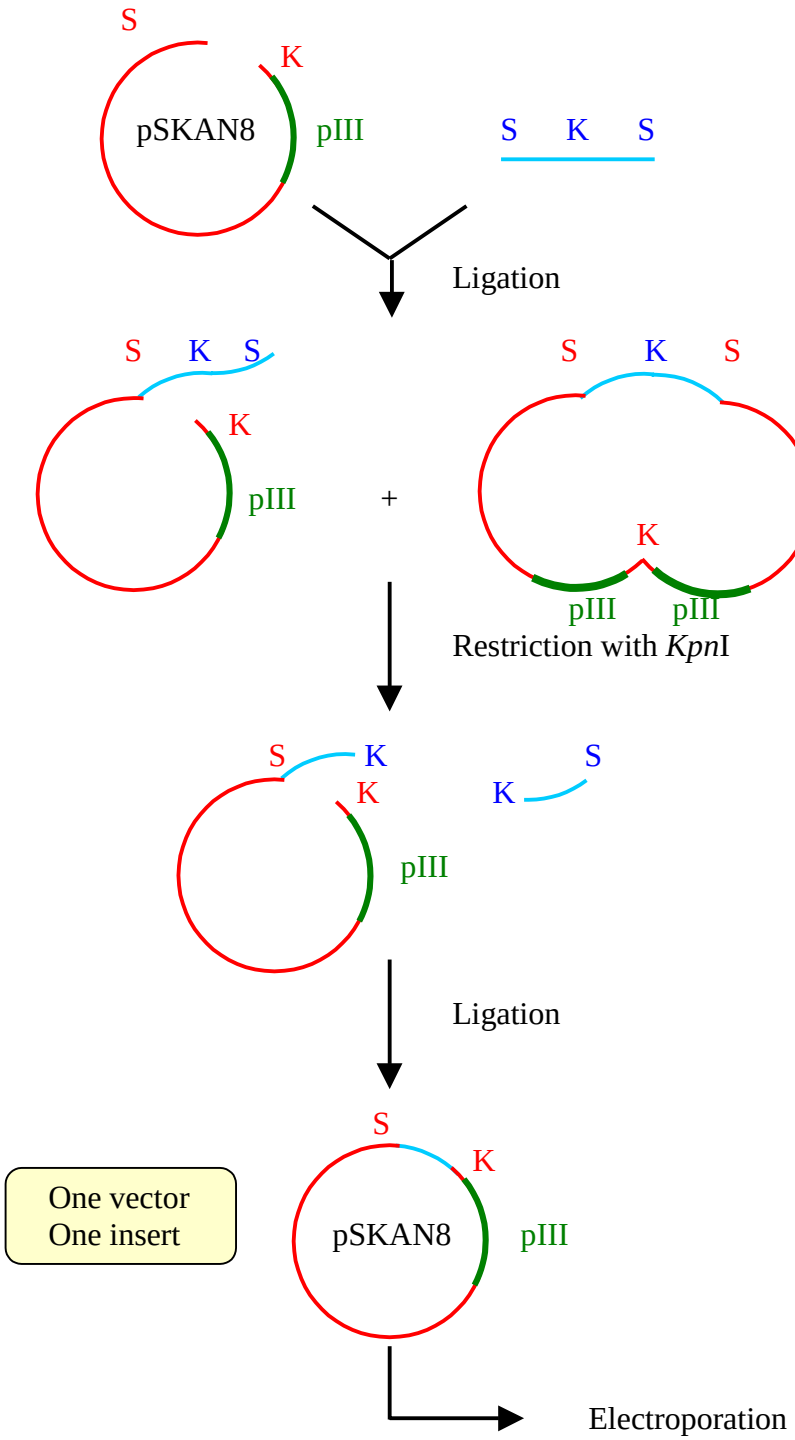


Figure III-7: Schematic representation of primary library construction. Opened vectors and dimeric inserts (prepared according to the protocol described in Figure III-6) are ligated together. Ligation between two vectors, in the absence of insert (formation of vector dimers) is prevented by using an excess of insert. Two vectors may bind to the same dimeric insert and thereby, blocking the *Kpn*I site, leading to a molecule composed of two vectors and one insert. Restriction with *Kpn*I exposes the *Kpn*I site of the insert which is then free to ligate with the second extremity of the vector, leading to ring closure. The ring closure is performed at low DNA concentration, to avoid the formation of concatemers. The vector with the fusion pIII protein can now be electroporated into bacteria, resulting in a high diversity library. S and K stand for *Sac*I and *Kpn*I sites, respectively.

Positions randomized (n)	Peptide diversity (20^n)	DNA diversity (NNK codons) (32^n)	Transformants required for complete library with: 90% confidence	99% confidence
1	20	32	74	149
2	400	$1.0x10^3$	$2.4x10^3$	$4.8x10^3$
3	$8.0x10^3$	$3.3x10^4$	$7.6x10^4$	$1.5x10^5$
4	$1.6x10^5$	$1.1x10^6$	$2.4x10^6$	$4.9x10^6$
5	$3.2x10^6$	$3.4x10^7$	$7.7x10^7$	$1.6x10^8$
6	$6.4x10^7$	$1.1x10^9$	$2.5x10^9$	$5.0x10^9$
7	$1.3x10^9$	$3.4x10^{10}$	$7.9x10^{10}$	$1.6x10^{11}$
8	$2.6x10^{10}$	$1.1x10^{12}$	$2.5\ x10^{12}$	$5.1x10^{12}$

Table III-1: Necessary sizes of phage display libraries to ensure the presentation of all possible random peptides. NNK codons (where N is any nucleotide, and K is G or T) is commonly used for randomization in biological libraries. NNK encodes all the amino acids except two of the stop codons, reducing the stop codons occurrence by two thirds. The confidence that a library contains all possible amino acid sequences is calculated assuming a Poisson distribution (Lowman and Wells, 1991). Adapted from Clackson and Wells (1994).

Library	Single clones selected on ampicillin agar plates
pSKAN8-HypD	$9x10^9$
pSKAN8-HypE	$1.8x10^8$
pSKAN8-HypF	$1x10^9$
pSKAN8-HypG	$1.3x10^9$

Table III-2: Estimation of pSKAN8-HypD, -HypE, -HypF, -HypG diversities. The number of single clones was estimated from dilution series from each performed ligation, plated on ampicillin LB plates.

2- Evaluation of the quality of the banks

Randomly selected clones were picked and sequenced, to confirm the presence of a single insert within the PSTI and to check the quality of the banks. The frequency of clones which are submitted to a change in the open reading frame (insertion, deletion) or which contain stop codons has to be estimated. In addition, the frequency of codons for each amino acid has to be determined and compared with the expected values.

An initial analysis of the sequenced clones showed that between 72% (HypD) and 94%

(HypE) of the clones contained one insert within the PSTI gene which was correctly formed (Table III-3). These clones did not exhibit mutations within the fixed amino acids contained in the hypervariable region, or within the constant region flanking this domain. In addition, they did not show a change in the open reading frame, which would lead to phage particles lacking pIII fusion protein.

In HypD and HypF libraries, some clones were found which contained 3 inserts (20,4% and 6,7% respectively). However, no clone containing religated vector without insert DNA was found, in any of the libraries.

Between 5,9% (HypE) and 13,4% (HypF) of the analyzed clones contained point mutations, such as insertions, deletions or exchanges of 1-3 bp.

	pSKAN8-HypD	pSKAN8-HypE	pSKAN8-HypF	pSKAN8-HypG
total of sequenced clones	44 (100%)	34 (100%)	30 (100%)	30 (100%)
correct sequences	32 (72,7%)	32 (94,1%)	24 (80%)	27 (90%)
3 inserts within PSTI	9 (20,4%)	0	2 (6,7%)	0
1bp insertion	1 (2,3%)	0	2 (6,7%)	1 (3,3%)
1-3bp deletion	2 (4,6%)	2 (5,9%)	2 (6,7%)	1 (3,3%)
1bp point mutation	0	0	0	1 (3,3%)

Table III-3: DNA sequence analysis of randomly selected clones from pSKAN8-HypD, -HypE, -HypF, -HypG. Relative frequency of correctly and incorrectly formed clones.

The correctly formed clones were submitted to further analysis. The hypervariable regions were further investigated to determine the ratio of particular bases at each of the three positions of the hypervariable codons (Table III-4).

The codon scheme used to encode the random amino acid sequences of the HypD, E, F, and G libraries, is NNK. With this NNK system, one would have expected to find 25% of each base in the first two positions of the codon and a frequency of 50% each, for G and T, in the third position. These theoretical values were respected for the third position of the codon, however, not for the first two positions, where G was dominantly found, to the detriment of A and C which were under-represented. T was also slightly over-represented at the first two positions of the codon.

In both HypD and HypF libraries, a very low frequency of A and C (0,5 to 1,1%, respectively) at the third position of the codon (“not allowed” position) was found. This may

have been due to the PCR parameters employed which were permissive and may have permitted a low frequency of Taq polymerase “errors”. These “errors” may have increased the diversity of the hypervariable sequences.

Bank	Base	Codon position		
		first	second	third
HypD	**A**	18	13,3	0,5
	C	17	20	0,5
	G	35,6	39,9	50
	T	29,3	26,6	48,9
HypE	**A**	15,9	17,1	0
	C	17,7	15,2	0
	G	35,4	35,4	48,2
	T	31	32,3	51,8
HypF	**A**	21,1	22,2	1,1
	C	20	17,8	1,1
	G	35,5	32,2	57,7
	T	23,3	27,8	40
HypG	**A**	10,8	19,3	0
	C	19,3	16,2	0
	G	40,8	38,7	54,8
	T	29	25,8	45,2
HypDEFG	**A**	16,5	18	0,4
	C	18,5	17,3	0,4
	G	36,8	36,5	52,6
	T	28,1	28,1	46,5

Table III-4: Percent base distribution within the codons of the hypervariable region of the pSKAN8-HypD, -HypE, -HypF, -HypG banks. The values of the HypD and HypE banks were calculated from 32 clones each (160 random codons). For HypF, and HypG, the values were calculated from 24 and 27 clones (144 and 162 random codons), respectively. Since the codon scheme used is NNK, where K is G or T, 25% of A, C, G and T were expected within the first two positions, and 50% G and T at the third position.

Each of the four libraries exhibited the same base position trends (Table III-4). Therefore, the base distribution of the four banks together was determined, and is shown in Table III-4.

The over-representation of G and the under-representation of A and C (Table III-4), allow a first estimation of the amino acid frequency that might be found for the different libraries. Codons rich in G, might be over-represented, whereas the ones containing several A

or C might occur at a very low level. Hence, one can predict that Gly (GGN), Val (GTN), but also Phe (TTT; TTC), Trp (TGG) and Cys (TGC; TGT) might appear more often than expected, based on the NNK theoretical values. On the other hand, Thr (ACN), Asn (AAT; AAC), Lys (AAA, AAG), Pro (CCN), His (CAT, CAC) and Gln (CAA, CAG) might be very rare.

	NNK	HypD	HypE	HypF	HypG	HypDEFG
Ala	6,2	6,5	3,8	1,1	9	5,1
Arg	9,3	9,8	11,4	7,7	7,8	9,2
Asn	3,1	0,5	1,9	2,2	1,1	1,4
Asp	3,1	2,2	3,1	5,5	4,5	3,8
Cys	3,1	4,9	3,8	2,2	6,7	4,4
Glu	3,1	2,7	3,8	3,3	2,2	3
Gln	3,1	1	1,9	1,1	3,3	1,8
Gly	6,2	16,8	11,4	17,7	18	16
His	3,1	1,6	0	3,3	3,3	2
Ile	3,1	2,2	3,8	2,2	1,1	2,3
Leu	9,3	6,5	7	7,7	7,8	7,2
Lys	3,1	0,5	1,3	1,1	1,1	1
Met	3,1	3,2	1,9	6,6	1,1	3,2
Phe	3,1	7,6	6,3	2,2	5,6	5,4
Pro	6,2	4,3	2,5	6,6	2,2	3,9
Ser	9,3	10,8	7,6	5,5	5,6	6,2
Thr	6,2	3,2	3,1	5,5	1,1	3,2
Trp	3,1	4,9	7	4,4	4,5	5,2
Tyr	3,1	1,6	2,5	2,2	1,1	1,8
Val	6,2	7,6	13,3	7,7	9	9,4
Stop codon	3,1	1	2,5	3,3	3,3	2,5

Table III-5: Expected and actual amino acid frequencies (in percentage) of the pSKAN8-HypD, -HypE, -HypF, and -HypG libraries. The last raw corresponds to the amino acid frequencies of the four banks taken together. This represents the data of 115 clones (626 random amino acids).

The actual amino acid frequency found in each of these libraries was determined from the correct clones which were sequenced. The results of this analysis were compared with the theoretical frequencies expected in a NNK library, and are summarized in Table III-5.

Figure III-8 gives an overview of the data presented in Table III-5. The amino acid frequency calculated for the four banks were correlated with the expected amino acid frequency of a NNK bank, however, it corresponded to the predictions based on the results of

the base distribution analysis (Table III-4): over-expression of Gly, Val, Trp, Phe and Cys, under-expression of Thr, Asn, Lys, Pro, His, Gln. With the exception of the histidine in HypE, every amino acid were represented in each bank.

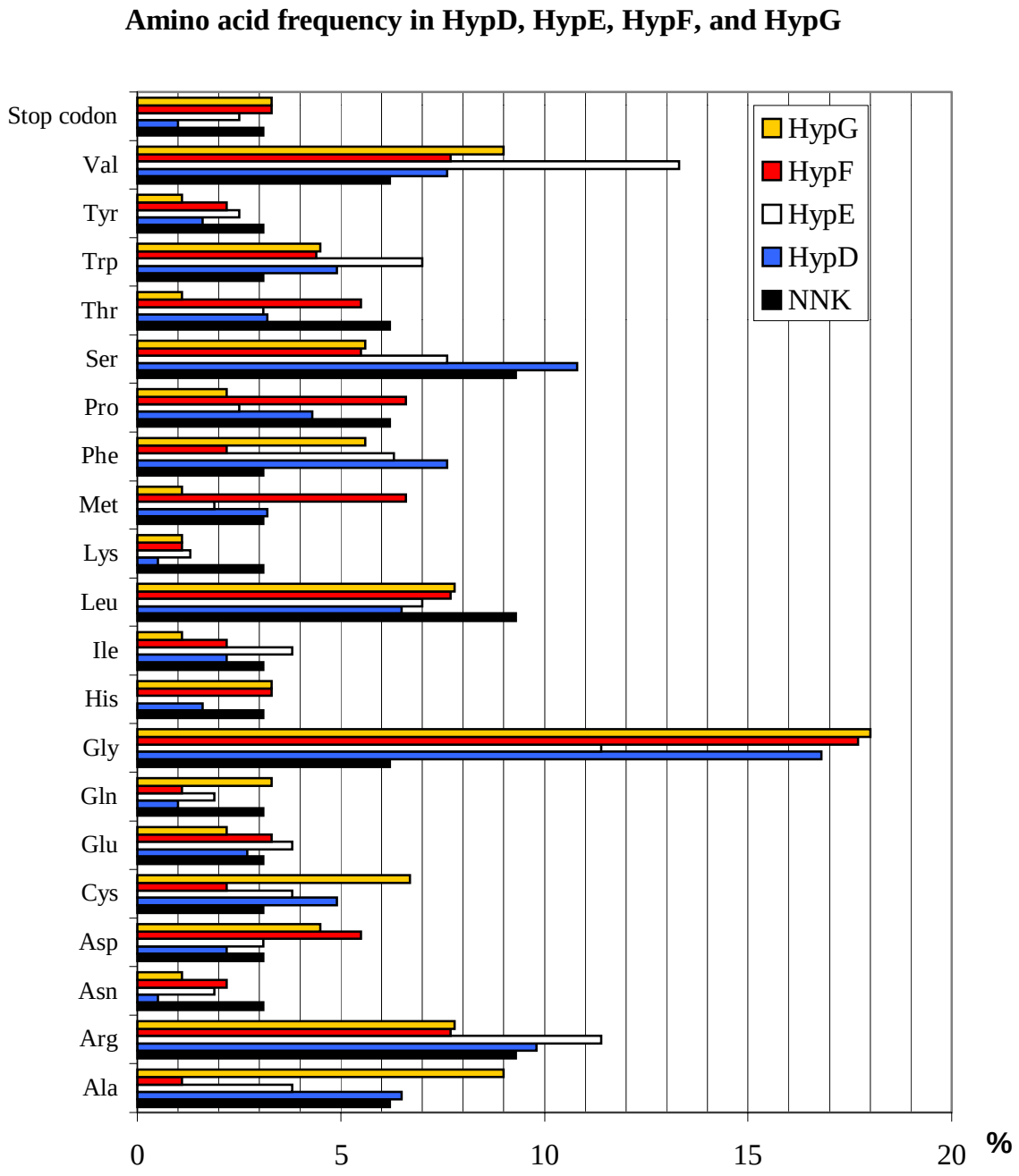


Figure III-8: Expected and actual amino acid frequencies (in percentage) of the pSKAN8-HypD, -HypE, -HypF, and -HypG libraries, according to the values presented in Table III-5.

3- Panning selection on SH3 domains

The quality of the four constrained random peptide libraries has been shown to be relevant in isolating peptides against a particular target. Indeed, the diversity of the banks was above the theoretical size required for pSKAN8-HypD and -HypE, and was almost as required for the two other banks, pSKAN8-HypF and -HypG (see Table III-1 and III-2). Furthermore, the libraries represent all codons, except the two forbidden codon stops. Therefore, the quality of the banks were considered sufficient to be tested against SH3 domain targets, in the attempt to isolate strong ligands.

In addition, the method used to prepare the SH3 domains as GST fusion proteins, was proven successful as demonstrated by SDS PAGE and Western blot. Most researchers who have previously worked with SH3 domains, have successfully used GST-SH3 fusion proteins (Rickles et al., 1994; Yu et al., 1994; Feng et al.,1995; Alexandropoulos et al., 1995; Bunnell et al, 1996; Sparks et al., 1996), suggesting that GST-SH3 fusion proteins are suited to panning selection of phage display libraries. Ligands for GST, when using GST-fusion proteins in affinity selection, have not been reported.

Although target proteins and libraries were correctly prepared, the panning selections were unsuccessful. Extended tests were performed on Src SH3 domain, all of which were negative. To ensure that the problems encountered were not dependent upon the target protein itself, Grb2 SH3 domain was also tested, without amelioration of the results.

The panning procedure was performed essentially as previously described, with some modifications, aimed at improving the interaction between the target protein and the phage library: the fusion proteins (contained in PBS) were immobilized using two types of supports, a maxisorb (Nunc), or a glutathione microtiter plate (Pierce), by incubation overnight at 4°C. In case of glutathione plates, the GST is captured by the glutathione fixed to the bottom of the wells, and the fused protein (Src) is expected to be orientated such that it is available for interactions with the phage library. However, no improvement was evident by changing the microtiter plate.

After immobilization, unspecific binding of the phage to the walls of the wells was blocked with PBS/2% skim milk powder, and incubations with the phage library subsequently carried out at room temperature for 2 hours. Previously, the library was diluted 1:1 in PBS/2% skim milk powder, to minimize the background.

After incubation of the target proteins with the phage library, wells are washed to

remove unbound, or low binding phage. The stringency of the washing steps increased with the cycles of selection, as shown on Table III-6.

On the first panning performed, PBS/0,5% Tween 20 was used to wash away the unwanted phage. As no enrichment of particular clones against the target protein was observed, PBS with 0,05% Tween 20 was used in subsequent wash steps. In addition, it was attempted to wash only with PBS during the first round of panning, and to introduce Tween only from the second cycle on. Reduction or elimination of the Tween during the first panning round might help to rescue clones which might be able to bind to the target, but which are under-represented in the library and, therefore, more easily lost during the washing steps. If such a clone is recovered after the first round of panning, it would be amplified (after elution, and reinfection) and be present in larger amounts in the second cycle, and Tween can be introduced. Unfortunately, none of these approaches led to the characterization of ligands for Src or Grb2 SH3 domains.

	cycle of panning		
	1	2	3 - 5
PBS/Tween	1x 5 min	3x 5 min	5x 5 min
PBS/milk	10 min	10 min	10 min
PBS/Tween	1x 5 min	3x 5 min	5x 5 min
PBS/milk	5 min	5 min	5 min
PBS/Tween	1x 5 min	3x 5 min	5x 5 min
H_2O	1x	1x	1x

Table III-6: Stringency of the washing steps according to the cycle of panning. The PBS/Tween contains the appropriate amount of Tween, according to the target protein, and the aim to reach. In our experiments, we used 0,5 or 0,05% of Tween 20. The PBS/2% skim milk powder was the blocking solution used. In case that some phage bound to the milk powder used to block the walls of the well, the milk added in solution competes this interaction.

Figure III-9 shows the evolution of the input (amount of phage incubated with the target protein), output (eluted phage after stringent washing steps), and the ratio of output/input (O/I) which is an indication of enrichment. When most of the added phage (input) is recovered after washing and elution (output), this indicates that the majority of the phage clones had a high affinity for the target, and remained attached to the target during the washing steps. These clones were enriched during the previous cycles of panning selection. The enrichment factor is estimated by dividing the ratio O/I from the round of interest (where

the output and the ratio O/I appear to be very high), with the ratio O/I from the first round. An enrichment is considered significant when the factor increases by approximately 100 or greater. It is impossible to recover all of the input, and a minimum background level cannot be avoided.

As seen in Figure III-9, no enrichment was observed. Furthermore, the titration values of pSKAN8-HypD on Src SH3 domain were very close to the values obtained for the negative control, where the library was incubated in a well, without immobilized target protein.

The titration values show in the Figure III-9 are for the phage library pSKAN8-HypD, tested by affinity selection with Src SH3 domain. However, the four banks, pSKAN8-HypD, -HypE, -HypF, -HypG, showed the same titration trends, when tested against Src, or Grb2 SH3 domains. The variations introduced to the panning protocol (different way to immobilize the target, or altering the stringency of the washing steps) did not lead to any improvement of the affinity selection.

No enrichment was evident after five rounds of panning, nevertheless, some randomly selected clones were picked and sequenced, to try to identify a potential homology between the clones. The analysis of the sequences shows that there was no similarity between the clones. It was obvious that no particular phage was retained by the target, since the clones which are amplified along the cycles of panning were the unproductive clones, due to mutations or stop codons. These smaller particles have a selective advantage, due to their shorter replication time.

Previous analysis of the original pSKAN8-HypD library demonstrated that less than 7% of the random peptides contained insertions or mutations (Table III-3), and only 1% of the amino acids were estimated to represent the TAG stop codon (Table III-5). If we consider that the pSKAN8-HypD library contains 5 random amino acids, only one stop codon is expected every 20 random peptides (5% of the random peptides contain a stop codon). That means that the original pSKAN8-HypD library contained approximately 12 % of unproductive clones.

As shown in Table III-7 and Figure III-10, the unproductive clones have been dramatically enriched during the panning, since they represent 28,6% of the population after the 2^{nd} round of panning, and dominate after the 4^{th} round (53,8%).

A

		Blank	Src
1st round	Input	3,10E+12	3,10E+12
	Output	3,70E+06	7,90E+05
	O/I	1,20E-06	2,50E-07
2nd round	Input	2,80E+11	2,40E+11
	Output	4,00E+03	3,50E+04
	O/I	1,40E-08	1,40E-07
3rd round	Input	4,30E+10	2,60E+10
	Output	9,50E+02	1,60E+03
	O/I	2,20E-08	6,10E-08
4th round	Input	5,00E+11	4,00E+11
	Output	3,00E+02	2,10E+03
	O/I	4,80E-09	4,50E-08
5th round	Input	4,10E+12	3,50E+12
	Output	2,80E+05	3,80E+05
	O/I	6,80E-08	1,00E-07

Panning of pSKAN8-HypD on Src SH3 domain

B

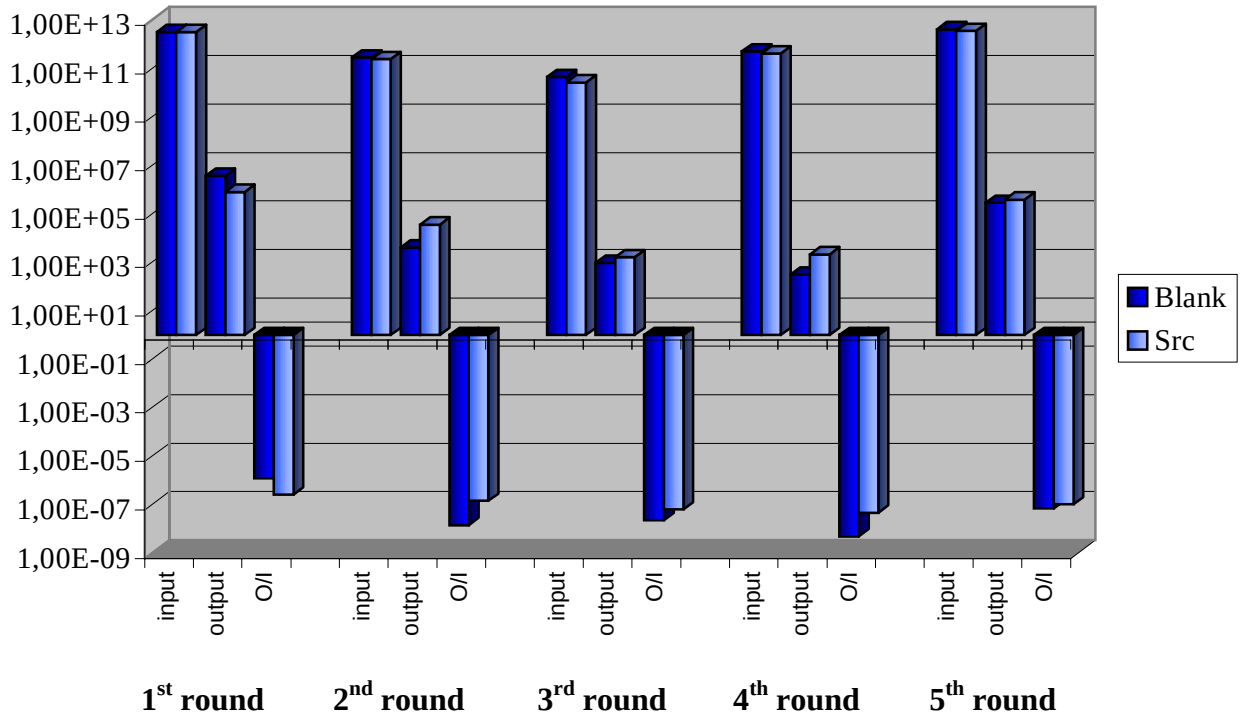


Figure III-9: Titration of the input, output and ratio output/input of the pSKAN8-HypD library, after 5 rounds of panning on Src SH3 domain. (A-) Titration values, in absolute number (cfu) of the input, output and ratio output/input, followed over 5 rounds of panning. (B-) Histogram of the values. Src was immobilized and the phage library was incubated with the target. A negative control was performed, where no target protein has been immobilized. It was submitted to the same treatment.

	2nd round	3rd round	4th round
correct sequence	71,1	55	46,2
unproductive clone	28,6	45	53,8
stop codon	21,5	25	30,7
insertion/deletion	7,4	20	23

Table III-7: DNA sequence analysis of randomly selected clones from pSKAN8-HypD after the 2nd, 3rd and 4th round of panning selection on Src SH3 domain. Relative frequency of correctly and incorrectly formed clones, in percentage. The incorrectly formed clones are unproductive clones, due to the presence of a stop codon, or a change in the reading frame, owing to insertions or deletions. 14, 60 and 26 clones were sequenced after the 2nd, 3rd, and 4th round of panning, respectively.

Figure III-10 shows the proportion of unproductive clones compared to correctly formed ones, and specify the frequency of stop codons, or mutations responsible for these malformed clones. Even if the trends tends to equilibrate after a few rounds, the highest cause of incorrect clones is the presence of stop codon(s) within the hypervariable domain.

The proportion of correct clones compared to unproductive particles for the pSKAN8-HypD library selected on Src SH3 domain (Figure III-10), shows high homology with the ratio of correct to incorrect clones, estimated for randomly picked clones after different rounds of panning selection of libraries pSKAN8-HypE, -HypF and -HypG, using Src, or Grb2 SH3 domains as targets.

This unsuccessful panning was disappointing, but finally, not really surprising. The possible reasons for expecting negative results with constrained peptide libraries are argued in the discussion.

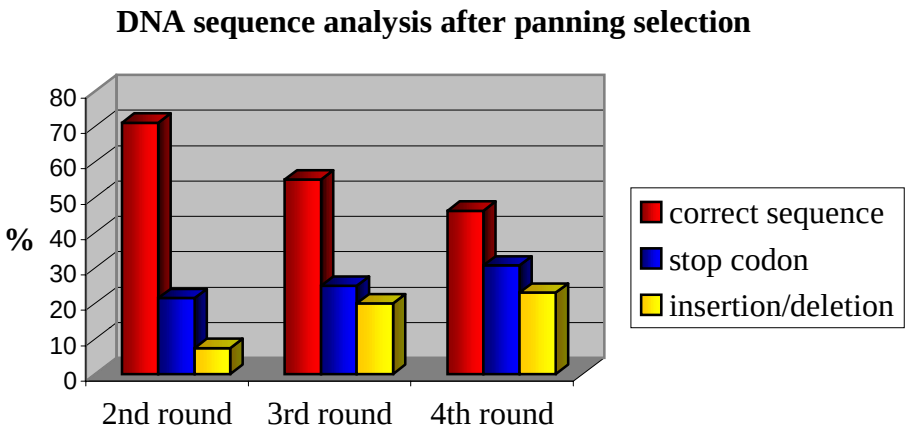


Figure III-10: Ratio of corrected formed clones to malformed particles after panning selection. DNA sequence analysis of randomly selected clones from pSKAN8-HypD after the 2nd, 3rd and 4th round of panning selection on Src SH3 domain. 14, 60 and 26 clones were sequenced after the 2nd, 3rd, and 4th round of panning, respectively.

4- SH3 ligands displayed on bacteriophage

To better understand why the panning selection performed with the constrained libraries failed, extensive tests were performed: SH3 ligand was inserted into the PSTI region of the pSKAN8 vector, at sites where the hypervariable domains were located and the four constrained libraries generated. These clones were used to test the hypothesis that two cysteines at both extremities of the variable region, may alter PPII conformation and thereby, obstruct the interaction between the SH3 domains and potential ligands.

A SH3 ligand was inserted into pSKAN8 following the protocol used to clone the hypervariable region, to create the constrained random peptide libraries. Our choice of SH3 ligand was based on the results of previous studies by other groups. The group of Brian Kay has worked extensively on SH3 domains, including isolation of ligands with distinct preferences for different SH3 domains (Src, Abl, Cortactin, p53bp2, Crk and Grb2) by using a phage display library, presenting the peptide X_6PXXPX_6 (Sparks et al., 1996). Based on this work, a ligand was selected, which was originally isolated against the N-terminal SH3 domain of Grb2 (Grb2N), which in phage ELISA, also shows strong cross-reaction with other SH3 domains, including Src, Yes, and NckN (N-terminal SH3 of Nck).

Oligonucleotides were designed to introduce the *Sac*I and *Kpn*I restriction sites at the extremities of the ligand, using PCR, and to facilitate cloning of the ligand into pSKAN8 (protocol was as described in Figures III-6 and -7).

Two forms of the ligand were prepared. Firstly a “short” ligand which had the size of the longest of our constrained libraries (pSKAN8-HypG). This short ligand conserves the minimum sequence required for specific binding to the SH3 domains (Arg Pro Leu Pro Pro Leu Pro). Two amino acids (Tyr Gln, on the N-terminal part), were introduced to lengthen to 9 amino acids within the two cysteines. The proline core motif (Pro Pro Leu Pro) is essential for binding of the ligand to the SH3 domains, while the N-terminal Arg Pro Leu motif confers specificity for Src SH3 domain (Sparks et al., 1994; Rickles et al., 1994;Rickles et al., 1995; Sparks et al., 1996).

A “long” ligand was also designed to test the influence of the size of the peptide on the constraint of the loop. This ligand contained the complete 18 amino acid sequence of the original Grb2N ligand. Since the loop is only partially constrained by the disulfide bridge, the amino acid sequence flanked by the cysteines may form additional constraints. The larger the loop, the less the constraint. Hence, a 18 amino acid loop may allow formation of the required PPII helix.

Original Grb2N ligand

Tyr **Tyr Gln Arg Pro Leu Pro Pro Leu Pro** Leu Ser His Phe Glu Ser

Short Grb2N ligand as displayed on the surface of the phage

Cys **Tyr Gln Arg Pro Leu Pro Pro Leu Pro** Cys

Long Grb2N ligand as displayed on the surface of the phage

Cys Pro Tyr **Tyr Gln Arg Pro Leu Pro Pro Leu Pro** Leu Ser His Phe Glu Ser Val Cys

The short and long ligands were cloned into pSKAN8 and electroporated in a *E.coli* strain, before being packaged by superinfection with the helper phage. The ready particles were tested on phage ELISA against Src, Grb2, Grb2N, and NckN SH3 domains. Neither the short nor the long ligands have been able to interact with the SH3 domains. A positive control was performed to ensure that the phage ELISA assay was functioning: one of the ligands isolated for chemotrypsin using this PSTI pSKAN8 vector system (Röttgen, 1996) was able to strongly react with chemotrypsin.

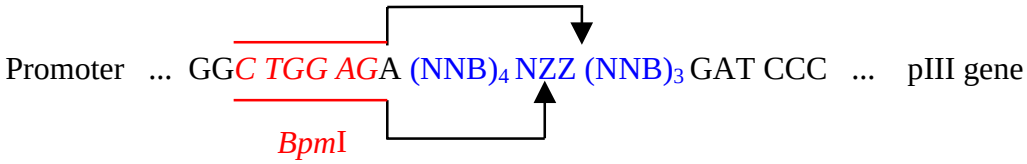
Taken together, our data suggest that SH3 ligands bracketed by cysteines, are not able to interact with SH3 domains, likely because they are not able to form the required PPII helix. However, a positive control where the original Grb2N ligand would have been cloned into pSKAN8 was missing, to make a complete demonstration of this hypothesis.

C – Cosmix-plexing® library: CPLPPXP

The second approach used to isolate strong ligands for SH3 domains was through the use of a cosmix-plexing® library. The cosmix-library generated to study SH3 and EVH1 domains was also a biased library (Figure III-3), with three fixed prolines, but with the variant peptides displayed as a linear prolongation of the pIII phage protein.

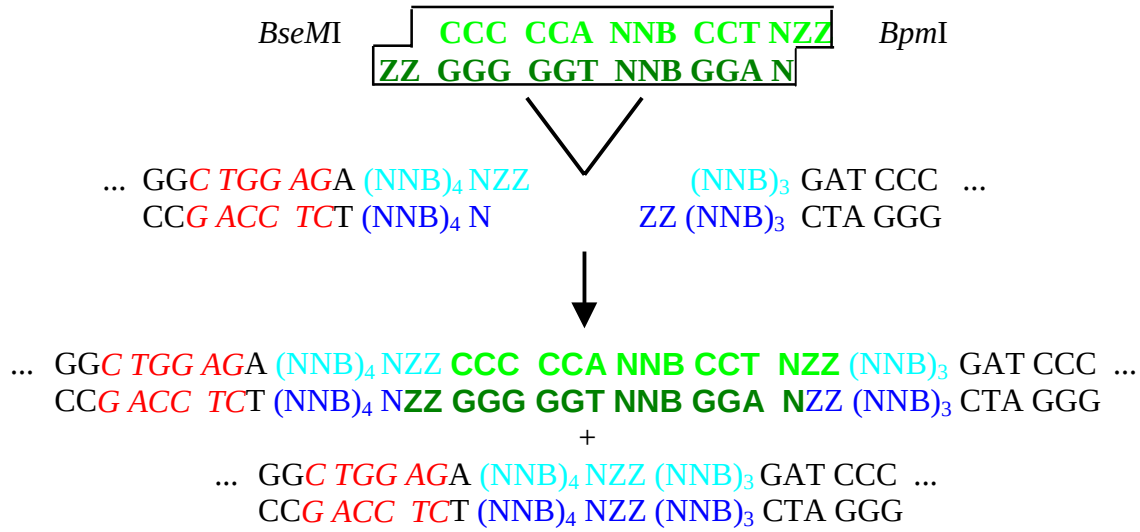
The cosmix-plexing® library was constructed in two steps, as illustrated in Figure III-11. First, the synthetic hypervariable oligonucleotides (24 bases, encoding 8 amino acids) was inserted into pROCOS4/7-stuffer1 (Figure III-11/A). The second step was to introduce the proline-rich project specific cassette into the previous 8 amino acid peptide insert (Figure III-11/B).

A- Primary library - DNA sequence



B- Insertion of the project specific cassette

DNA sequence:



Amino acid sequences:

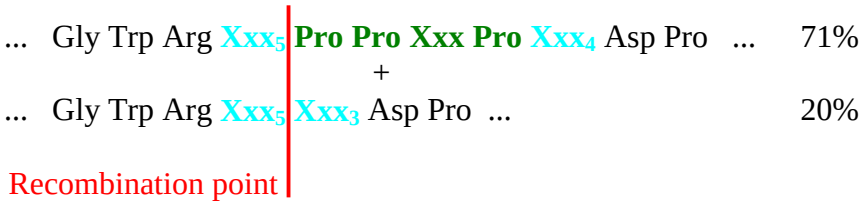


Figure III-11: Principle of the strategy of the CPLPPXP library construction. (A-) The primary library was generated by insertion of a completely random region, representing a 8 amino acid peptide. (B-) Due to the type IIs restriction enzyme *Bpm*I, the hypervariable fragment can be opened in the middle, and the proline rich cassette introduced. The cassette was digested at its extremities with *BseM*I and *Bpm*I type IIs restriction enzymes. When the cassette is ligated within the hypervariable domain, a proportion of the vectors may religated without cassette. These parental clones were estimated as 20% of the whole library. 9% of the library were incorrectly formed. The recombination point, where cosmix-plexing® induced recombination takes place, used to recombine left and right parts of the peptides, is at the end of the codon for the fifth random amino acid. ZZ represents AM (AA, AC), CT, GG, TS (TC, TG).

The name given to this final library is the CPLPPXP library. CPL stands for Cosmix-Plexing Linear peptide, and the PPXP represents the proline rich cassette motif (Pro Pro Xxx Pro). In the CPLPPXP library, **$(Xxx)_5$ Pro Pro Xxx Pro $(Xxx)_4$** variant peptides are displayed.

The cosmix-plexing® recombination can be carried out on this extension library by cleavage at the type IIs restriction site *(Bpm*I), allowing the optimization of the whole region. The *Bpm*I recognition site is located directly before the hypervariable domain, but *Bpm*I digests the DNA within the variant, after the fifth random codon (Figure III-11).

The phagemid pROCOS4/7-stuffer1 (Figure III-12) is an expression vector where variant peptides can be fused to the N-terminus of the pIII gene. pROCOS4/7, like all phagemid vectors, does not contain any phage genes, other than the pIII gene used as fusion partner. To be packaged into phage-like particles, pROCOS4/7 needs superinfection with a helper phage (M13KO7), which provides required bacteriophage functions.

In pROCOS4/7-stuffer1, the fusion-pIII hybrid synthesis is under the control of the λ P_L promoter. The λ P_L promoter is strongly repressed by the λ cI protein, which is produced in *E.coli* strains containing an integrated copy of the lambda genome (i.e. lysogens). The λ P_L promoter ensures that the correct ratio of hybrid to wild type pIII represents essentially monovalent presentation (i.e., one hybrid pIII per particle).

The “outer membrane”-protein leader peptide (Omp_L), ensures that secretion of the fusion peptide and correct cleavage of the leader peptide occur. pROCOS4/7-stuffer1 also contains plasmid replication functions, M13 phage replication/packaging origin, and the ampicillin resistance gene.

The stuffer is a 950 bp DNA fragment derived from pBR322. Since two different restriction enzymes are used for the insertion of the peptide to be fused to pIII, the stuffer (bracketed by the two restriction sites) allows to check for complete digestion of the vector, since the stuffer and the opened vector have different sizes and are, therefore, distinct on a gel.

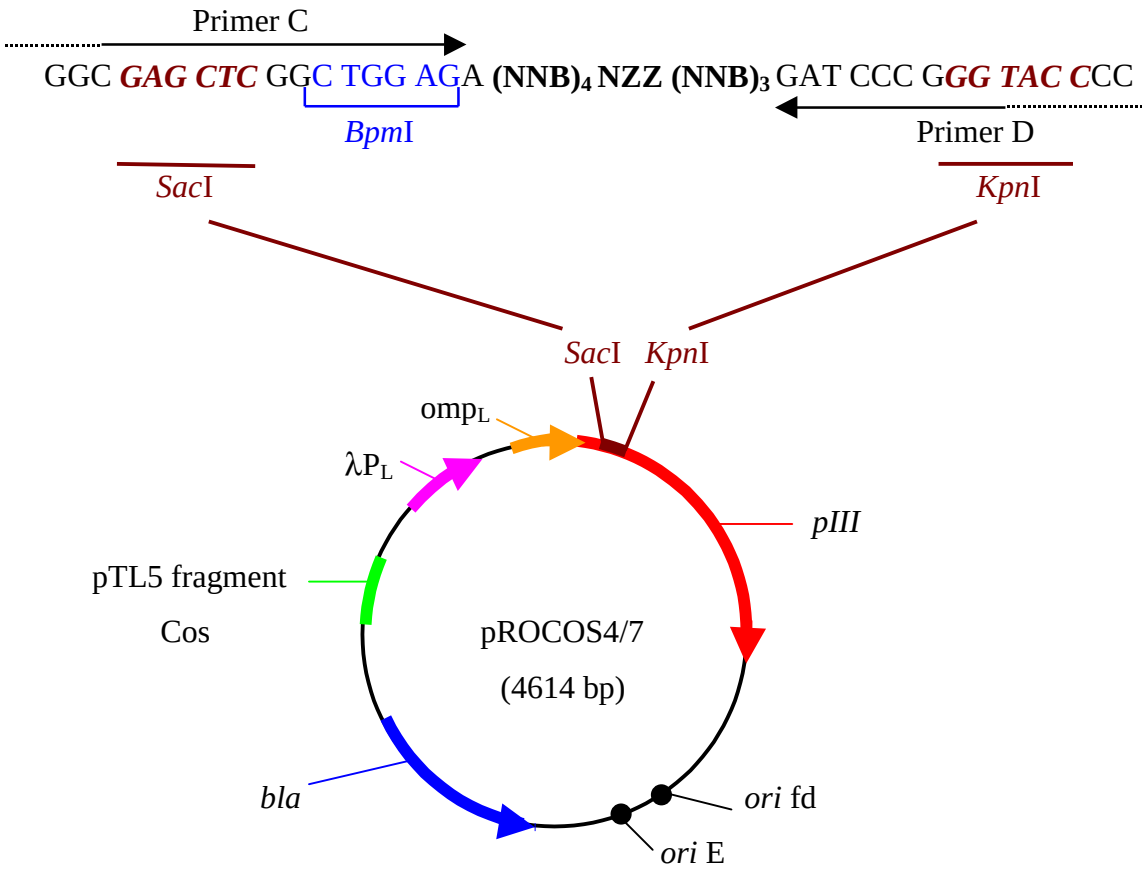


Figure III-12: Schematic map of the phagemid vector pROCOS4/7. The generation of the cosmix-plexing® library takes place in two steps. First, a hypervariable oligonucleotide is inserted into pROCOS, N terminal from pIII. Then, a project specific cassette is introduced within the variant region, due to a type IIs restriction enzyme (*Bpm*I), which digests a fixed distance away (within the hypervariable region) from the recognition site (within the flanking region). The endonuclease sites *Sac*I and *Kpn*I are used to insert the hypervariable oligonucleotides. The hypervariable inserts were amplified with the primers C and D (see section II-4-3-2), and digested with the restriction enzymes *Sac*I and *Kpn*I. The cassette was amplified with the primers C and E (see section II-4-4-2), and digested with the restriction enzymes *BseM*I and *Bpm*I, which give 2 bp cohesive ends, compatible with the digestion of the hypervariable region with *Bpm*I. In pROCOS4/7-stuffer1, the stuffer is bracketed by the *Sac*I and *Kpn*I restriction sites. The pTL5 fragment was added to the vector, in order to establish a cos site (necessary for λ packaging). The following abbreviations are used: λP_L: major leftward promoter of the bacteriophage λ; Omp_L: signal sequence derived from the E.coli outer membrane protein OmpA; *pIII*: gene coding for the minor coat protein pIII of the M13 bacteriophage; *bla*: β-lactamase gene (ampicillin resistance); *ori* fd: origin of replication for single stranded DNA of the bacteriophage fd; *ori* E: origin of replication for plasmids colE1.

1- Generation of the primary library: 8 amino acid random peptides are displayed

1-1 Preparation of the linearized vector DNA

The insertion site for the hypervariable oligonucleotide was prepared by linearizing the vector with the *Kpn*I and *Sac*I restriction enzymes. The hypervariable region was inserted between a leader "signal" sequence and the M13 *pIII* gene, so that the peptide was presented on the surface of the phagemid particle as a pIII amino terminal extension.

1-2 Preparation of the ds DNA insert from degenerated oligonucleotides

The codon scheme used to encode random amino acid sequences was NNB, where N is A,C,G or T, and B represents C, G or T. This scheme allows the reduction of the stop codon frequency (which would lead to unproductive clones), as only one of the three stop codons (TAG) can be encoded, while codons for all the 20 amino acid are still represented.

Four sets of inserts were prepared which led to four separate libraries. The four sets differed only by the dinucleotide corresponding to the cohesive ends generated upon cleavage by the type IIs restriction enzyme *Bpm*I (at the ZZ position, in Figures III-11 and III-12). The *Bpm*I type IIs restriction enzyme binds a defined sequence within the invariable sequence adjacent to the inserts, but cleaves at a fixed distance from the binding site, within the middle of the hypervariable domain, creating a 2 bp cohesive end. ZZ stands for AM (AA and AC), CT, GG and TS (TC and TG). It is important that no head to head joining of the cohesive ends occurs on religating the cleaved sequenced. This is achieved by the set of base pairs used at the ZZ position. Furthermore, when the four libraries were pooled, this NZZ allowed the presentation of all amino acids, except cysteine. The *Bpm*I restriction site is necessary for the insertion of the cassette and for the cosmix-plexing® recombination, whereas the *Sac*I and *Kpn*I restriction sites were used to create the original insert.

Although the strategy chosen to prepare the oligonucleotides involves many steps, it does guarantee pure products. This strategy is very similar to the one used to generate the constrained libraries and is illustrated in Figure III-6. The quality of the digested inserts have a strong influence on the quality of the ligation with the vector and, thus, on the efficiency of the electroporation of the ligated DNA into the bacteria. The constrained library protocol was optimized as shown in Figure III-13. The insert oligonucleotides were amplified via PCR, using biotinylated primers, leading to a product carrying biotin at the extremities. The difference comes from the digestion of the PCR products, which are cleaved by both *Kpn*I and

*Sac*I. The small restriction fragments are captured and removed with the superparamagnetic polysterene beads, covalently attached to streptavidin (Dynabeads® M-280 Streptavidin; Dynal). The inserts are only in this step, ligated among themselves in order to build a long chain of monomers (concatemer). With this protocol, it is possible to remove the small digested fragments earlier and to avoid that they eventually religate with the inserts to form unproductive clones. Therefore, the formation of the desired products is completely controllable.

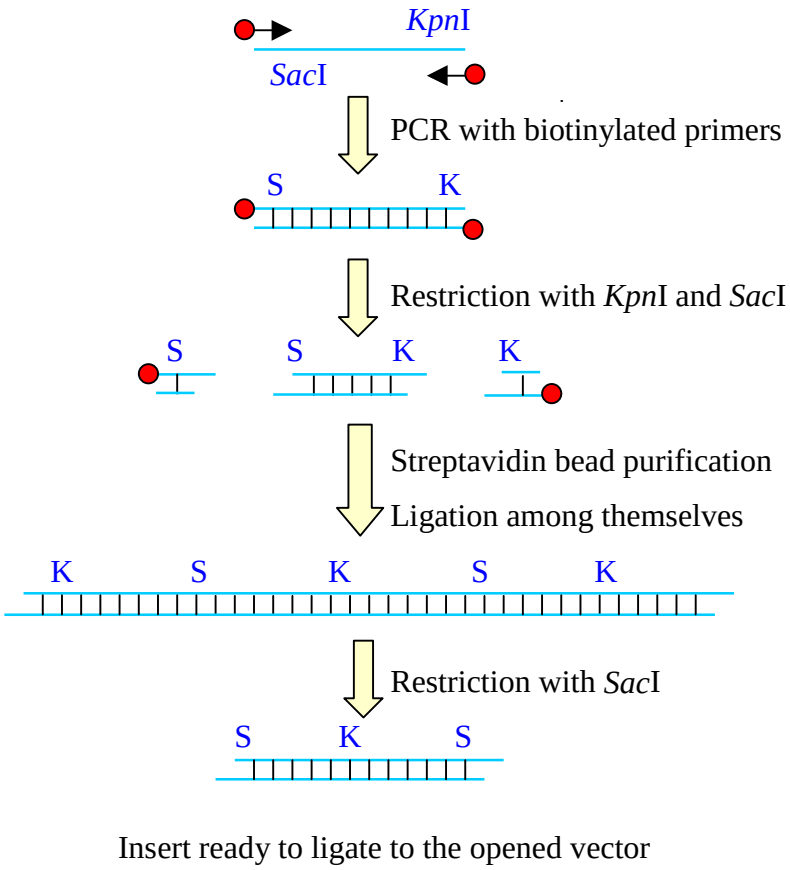


Figure III-13: Schematic representation of insert preparation. The insert oligonucleotides were amplified by PCR, using biotinylated primers, leading to a product carrying biotin on both extremities. The PCR product was cleaved with *Kpn*I, and *Sac*I. The small restriction fragments were removed by magnetic streptavidin beads. The inserts were then ligated among themselves. And resolved as dimeric inserts by digestion with *Sac*I. S and K stand for *Sac*I and *Kpn*I sites, respectively.

The DNA fragment is finally cleaved with *Sac*I and a dimeric insert is formed with a central *Kpn*I site and *Sac*I cohesive ends. The insert is ready to be ligated to the linearized vector.

1-3 Ligation of the vector and insert DNA

The digested vector (with *Sac*I and *Kpn*I ends) and inserts (which have *Sac*I cohesive ends) are ligated together. The same protocol as for the constrained libraries (as illustrated in Figure III-7) was used. The insert could only be ligated with one extremity of the vector, the

one digested with *Sac*I. Restriction with *Kpn*I was performed. The vector attached to one end of the insert was then ligated, leading to the ring closure of the DNA. This ligation was performed at low DNA concentration to encourage ring closure and discourage reinsertion of the small cleavage products.

The recombinant vectors thus formed, were electroporated into a WK6λmutS *E.coli* strain (Table III-8). The mixture of the four banks bring a total diversity of $2x10^8$ transformants. According to the Table III-1, one would need around $8x10^{10}$ variant to ensure that all possible codons are represented, with a 90% confidence level. However, since the clones will be recombined together by cosmix-plexing® when the cassette will be introduced, many new clones will be created, and the diversity is sufficient, i.e. all possibilities are present for both small 5 random amino acid blocks on each side of the cleavage site.

Library	single clones selected on ampicillin agar plates
PROCOS-AM	$6.5x10^7$
PROCOS-CT	$3.6x10^7$
PROCOS-GG	$5.7x10^7$
PROCOS-TS	$4.4x10^7$
Total pROCOS-insert	$2x10^8$

Table III-8: Estimation of pROCOS-AM, -CT, -GG, -TS diversities. The amount of single clones was estimated from dilution series from each performed ligation, plated on ampicillin LB plates.

2- Evaluation of the quality of the four primary libraries

It was important to ensure that the four primary libraries were well prepared, before inserting the proline rich cassette. Randomly selected clones were picked and sequenced, to confirm the presence of a single insert into the vector and to check the quality of the banks. The frequency of clones which underwent a change in the open reading frame (insertion, deletion) or which contained stop codons, was estimated. In addition, the frequency of codons for each amino acid was determined and compared with the expected values.

An initial analysis of the sequenced clones showed that around 80% of each of the four primary banks, contained one correctly formed insert within the vector (Table III-9). The other 20% of the clones led to unproductive phage particles, due to changes in the reading frame (insertions, or deletions of one or two bases) or the presence of a stop codon.

	pROCOS-AM	pROCOS-CT	pROCOS-GG	pROCOS-TS
total of sequenced clones	40 (100%)	26 (100%)	23 (100%)	27 (100%)
correct sequences	32 (80%)	21 (80,7%)	18 (78,2%)	22 (81,5%)
no insert	1 (2,5%)	0	0	0
1-2bp insertion	2 (5%)	2 (7,7%)	0	0
1-2bp deletion	1 (2,5%)	1 (3,8%)	2 (8,7%)	3 (11,1%)
Stop codon	4 (10%)	2 (7,7%)	3 (13%)	2 (7,4%)

Table III-9: DNA sequence analysis of randomly selected clones from pROCOS-AM, -CT, -GG, -TS. Relative frequency of correctly and incorrectly formed clones.

The correctly formed clones were submitted to further analysis. The hypervariable regions were further investigated, to determine the amino acid frequency in each of the bank (Table III-10). The results of this analysis were compared with the theoretical frequencies expected in a NNB library.

	NNB	AM	CT	GG	TS	Total
Ala	6	2,5	13,5	1,4	7,7	6,2
Arg	8	7,4	11,3	5,7	8,9	8,3
Asn	4	3,1	1,6	2,8	4,4	2,9
Asp	4	3,7	3,2	4,3	3,3	3,6
Cys	4	4,9	4,8	8,6	3,3	5,4
Glu	2	1,8	4,3	1,4	1,1	2,1
Gln	2	2,5	1,6	2,8	1,1	2
Gly	6	6,2	9,2	2,8	6,6	6,2
His	4	3,7	0,5	1,4	3,3	3,5
Ile	4	6,2	2,1	5,7	7,7	5,4
Leu	8	8,7	10,2	8,6	8,9	9,1
Lys	2	1,8	1	2,8	3,3	2,2
Met	2	2,5	3,2	2,8	3,3	3
Phe	4	4,3	1,6	11,4	6,6	2,9
Pro	6	7,4	3,8	5,7	5,5	5,6
Ser	10	11,1	10,2	4,3	10	8,9
Thr	6	5,6	3,8	5,7	2,2	4,3
Trp	2	1,8	3,8	4,3	1,1	3,3
Tyr	4	4,3	0,5	5,7	4,4	3,7
Val	6	6,8	9,2	12,8	5,5	8,5
Stop codon	2	1,8	1	1,4	1,1	1,3

Table III-10: Expected and actual amino acid frequencies (in percentage) of the pROCOS-AM, -CT, -GG, and -TS libraries. The last column correspond to the amino acid frequency of the four mixed banks. The calculations was determined from the correct clones which were sequenced: 32 clones for AM, 21 for CT, 18 for GG, and 22 for TS.

The Figure III-14 gives an overview of the data presented in Table III-10. The amino acid frequency calculated for the four banks correlate well with the expected values of a NNB bank. However, some variations were evident in the individual libraries after closer inspection.

Amino Acid frequency in pROCOS-AM, -CT, -GG, and -TS

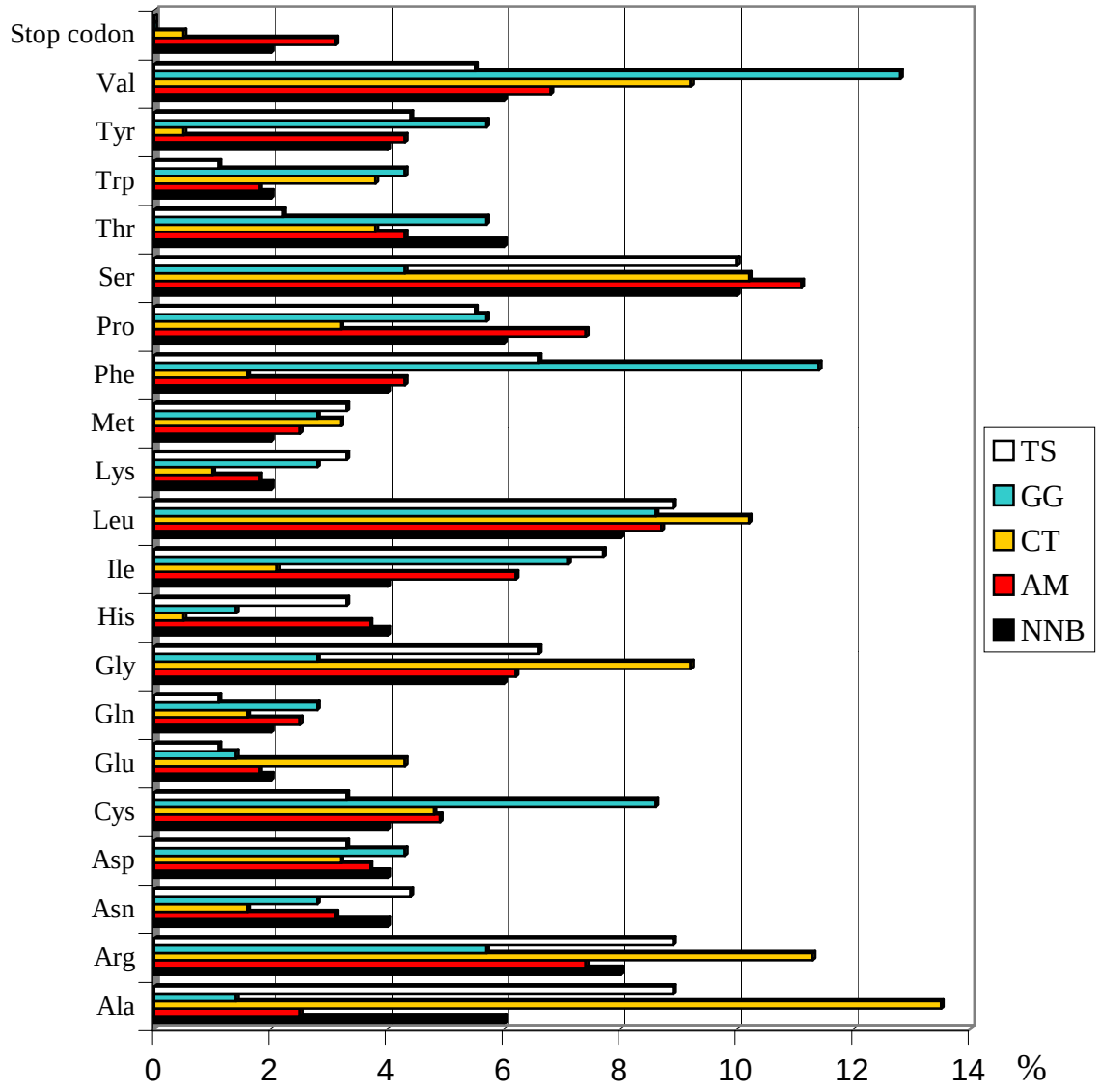


Figure III-14: Expected and actual amino acid frequencies (in percentage) of the pROCOS-AM, -CT, -GG, -TS libraries, according to the values presented on Table III-10.

The AM and TS libraries reflected the values according to the theoretical expectations, except a low frequency of alanine and threonine. CT was very weak concerning Asn, His, Phe, and Tyr, and too strong in Ala. As far as GG was concerned, it was deficient in Ala, Gly, His and Ser, whereas Cys, Phe, and Val were dramatically over-represented. Taken together, since the four banks were pooled for the insertion of the proline rich cassette, the amino acids were likely to be properly represented.

This analysis provides evidence that the four primary banks are capable of receiving the project specific cassette. The second step of the generation of the library was then performed: the introduction of the proline rich cassette.

3- Generation of the CPLPPXP library, insertion of the proline rich cassette

The next step was to insert this project-specific proline rich cassette into the hypervariable sequence created in the primary library, thus generating a secondary project-specific library. Invariable amino acid positions are chosen in the design of this cassette (Pro Pro Xxx Pro), so as to improve the interaction possibilities between this secondary library and the SH3 or EVH1 domains. The codon scheme used for the hypervariable codons, is NNB as it was for the initial inserts (B represents C, G or T).

16 sets of cassettes are prepared, which differed only by the two ZZ dinucleotides representing the cohesive ends generated on cleavage by *BseM*I and *Bpm*I, two type IIs restriction enzymes (Figure III-11/B). ZZ stands for AM (AA and AC), CT, GG, and TS (TC and TG). It is essential that the ZZ dinucleotides present on both cohesive extremities of the cassettes, differ on the right and left ends (i.e. whatever the ZZ set found on the right part of the cassette, AM, CT, GG and TS have an equal chance to be present on the left side). These distinct extremities promote the formation of concatemers during the cloning of the cassette into the primary library, as one cassette cannot be cloned within a single vector. This concatemer formation, strongly increases the diversity of the bank.

The sets used at the ZZ position prevented head to head joining of the cohesive ends during ligation of the cleaved sequences with the primary library. Furthermore, this set of dinucleotides allowed, when the 16 cassettes were pooled, the representation of all amino acids at the ZZ position, except cysteine, and reduced the frequency of the stop codons (only TGA is allowed).

*BseM*I and *Bpm*I were used to clone the cassettes within the primary library. Both generate a 2 bp cohesive end on both extremities of the cassettes, allowing a compatibility with the cohesive ends (set of 6) created by *Bpm*I cleavage of the initial library, within the hypervariable insert.

The optimized protocol for cloning the cassette into the insert of the primary library, is shown in Figures III-15, -16 and -17. Here too, a complex strategy was used, but which yielded extremely high electroporation efficiency (up to 10^8 transformants/µg DNA).

3-1 Preparation of the ds DNA proline-rich cassette from degenerated oligonucleotides

The cassette oligonucleotides are converted into double-stranded DNA and amplified via PCR, using biotinylated primers. They were restricted with *Bpm*I. These half opened cassettes were ready to be ligated with the vector (Figure III-15).

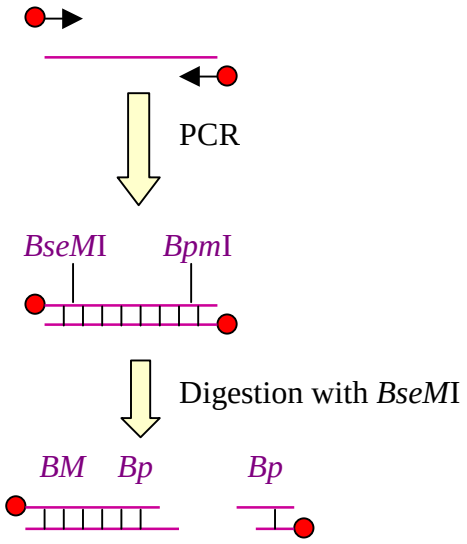


Figure III-15: Preparation of the double stranded project-specific cassette from degenerated oligonucleotides. The cassette is converted in double stranded DNA via PCR, using biotinylated primers. The cassette is then digested with *Bpm*I, and is ready to be ligated to the vector. *BM* and *Bp* stands for *BseM*I and *Bpm*I respectively.

3-2 Preparation of the linearized vector of the primary library

The vector containing the insert (the primary library) was used as cloning vector to create this extension library. It was first cleaved by *Bpm*I, opening the hypervariable region in the middle. Then, a second restriction with *BseR*I cut the vector into two parts, a small (S) and a large (L) fragment which were extracted from a gel and kept separate (Figure III-16).

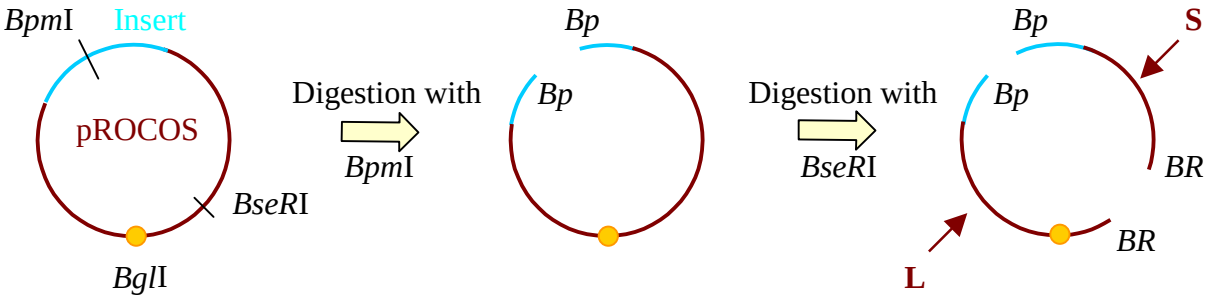


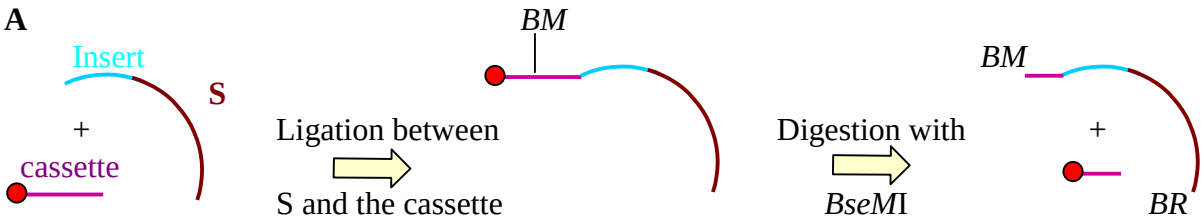
Figure III-16: Preparation of the linearized vector of the primary library. The vector is first cleaved within the hypervariable region by the *Bpm*I type IIs restriction enzyme. Then, digested a second time by *BseR*I, to cut the vector into two pieces, a small (S) and a large (L) fragment. *Bp* and *BR* stands for *Bpm*I and *BseR*I restriction enzymes, respectively. The restriction site of *Bgl*I is represented by an orange circle.

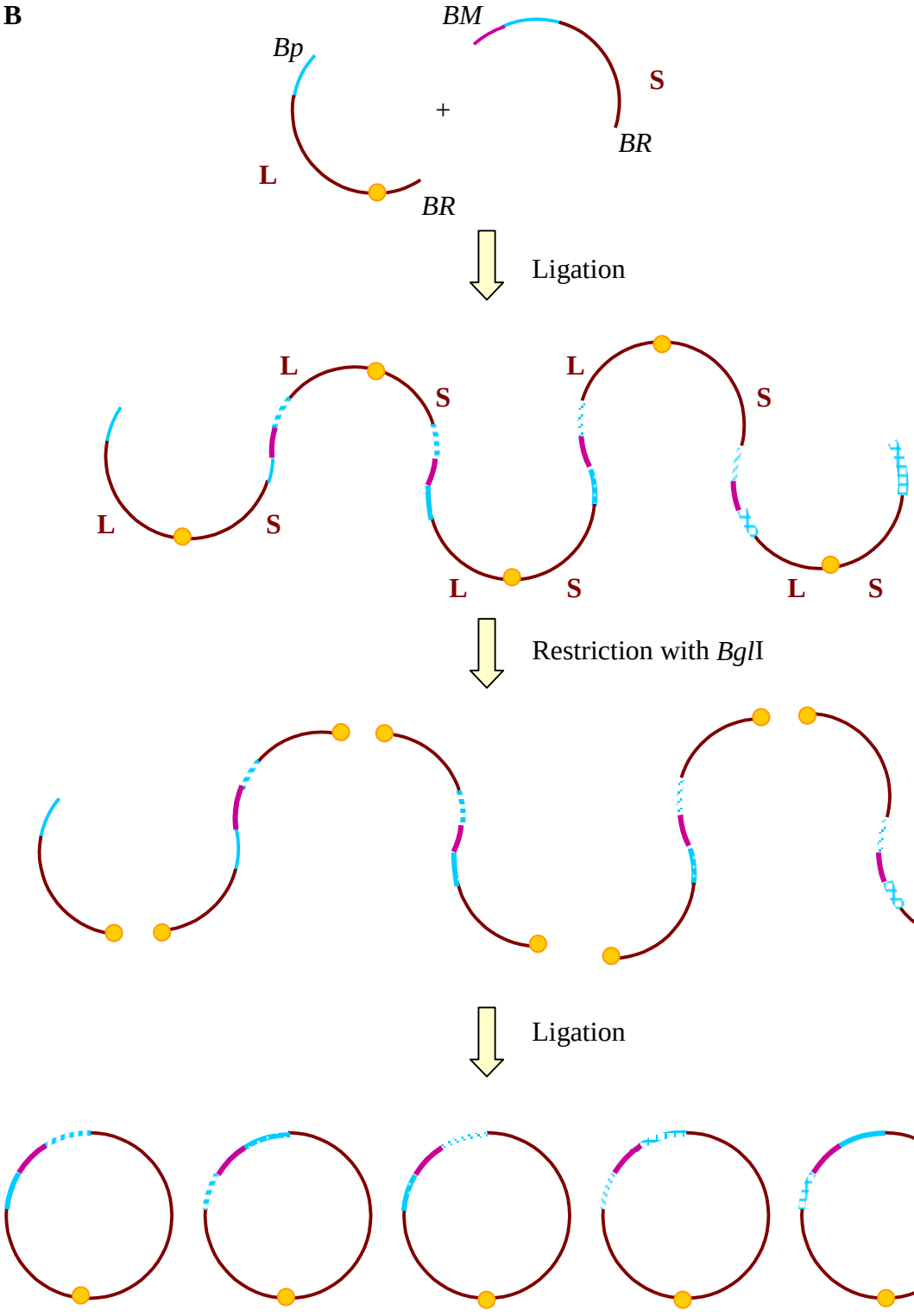
3-3 Insertion of the project-specific cassette into the primary library

The half opened cassettes were ligated with the small fragment (S) of the vector (S-cassette product), and cleaved with *BseMI* (Figure III-17). Unwanted ligation products and the small biotinylated ends from the cassette template were removed with streptavidin beads.

The large fragment (L) and the S-cassette, product from the preceding reaction, were ligated together at equimolar ratios and high DNA concentration (>200 µg/ml), reconstituting the complete vector, and leading to the formation of concatemers. Concatemers contained monomer units all oriented in the same direction, due to the fact that the restriction site of *BseR*I produces cohesive ends were not compatible with those produced at the *Bpm*I or *BseM*I cleavage sites. The ZZ set used did not allow head to head joining of fragments. As the two extremities of the cassettes do not present (in most cases) identical ZZ dinucleotides, and as such cannot be cloned into a single vector, most of the ligation products arise from fragments which come from different clones from the initial library. The ligation thus ensured a recombination among the vectors, leading to the formation of a more diverse hypervariable region.

Figure III-17: Insertion of the project-specific cassette within the primary library. (A-) The half opened cassette is ligated with S, and digested with *BseM*I. Restriction fragments are removed by streptavidin beads. (B-) Ligation occurs between L and the S-cassette product, at high DNA concentration, to direct the formation of concatemers. Concatemers are resolved as monomers by cleavage with *Bgl*I. Ligation at low concentration generate ring closure of the vector which now contain the hypervariable region complemented by the project-specific cassette. The recombined vector can be electroporated, packaged, and used in panning to isolate strong ligands against SH3 or EVH1 domains. *Bp*, *BM* and *BR* represent the *Bpm*I, *BseM*I and *BseR*I restriction sites, respectively. S and L stand for small and large vector fragments, respectively. The orange circle represents the *Bgl*I restriction site.


A
Insert
S
+
cassette
Ligation between
S and the cassette
BM
Digestion with
BseMI
BM
+
BR


B
BM
Bp
S
+
BR
L
BR
Ligation
L
S
L
S
L
S
L
S
L
S
Restriction with BglI
Ligation

The concatemers were resolved into single phagemid molecules on cleavage with *Bgl*I, followed by ligation at a low DNA concentration, to favor the formation of monomers.

This secondary library was then electroporated, packaged and transformed into WK6λmutS *E.coli* strain. This lambda lysogen strain expresses the cI transcription repressor of the lambda bacteriophage, and prevents an inopportune production of pIII gene product which might be toxic. In addition, it ensures the presentation of mainly one hybrid pIII per particle. DNA prepared in this way gives extremely high transformation frequencies (around 10^8 transformants/µg DNA).

Each vector contained a hypervariable region of 13 amino acids, of which three were fixed prolines (insert + cassette), as shown:

DNA sequence: $(NNB)_4$ NZZ CCC CCA NNB CCT NZZ $(NNB)_3$

***Peptide sequence*:** Xxx_4 Xxx Pro Pro Xxx Pro Xxx Xxx_3

The electroporation was carried out, until relevant bank diversity was reached. The CPLPPXP library contains 10 random residues and a diversity equivalent to $2{,}3x10^8$ transformants was obtained. This was obviously not enough to ensure the representation of all possible peptides, but, the diversity was increased each time a cosmix-plexing® was performed.

4- Evaluation of the quality of CPLPPXP library

The quality of the completed CPLPPXP library had to be evaluated.

35 randomly picked clones were sequenced, and revealed that around 71% of the library possessed the proline rich cassette (Table III-11); 20% of the clones were religated without the cassette and therefore, displayed only a 8 amino acid long peptide, meaning that 91% (71+20) of the bank was adequate for successful panning selections. The other 9% were non correctly formed particles.

When the vector was opened for the insertion of the cassette into the hypervariable region, it was cut in two pieces, the S and L fragments (Figure III-16). It has been shown that in particular cases, the L fragment is able to religate on its own. The *BseR*I digestion creates a 2 bp cohesive end, as shown in Figure III-18, which exposes a free AG dinucleotide at the L extremity. The 2 bp cohesive ends generated at the *Bpm*I restriction site present 6 variable

dinucleotides, according to the set of insert cloned in the vector (AA, AC, CT, GG, TC, and TG). The two extremities of the L fragment are able to interact together, when the TC set of insert is revealed at the *Bpm*I site. That means that one sixth of the L fragment are able to religate, even in the absence of the S fragment. By altering the ratio of cassette, S and L fragments, it was possible to reduce the formation of L clones to a minimum, however, they still represented 9% of the final library.

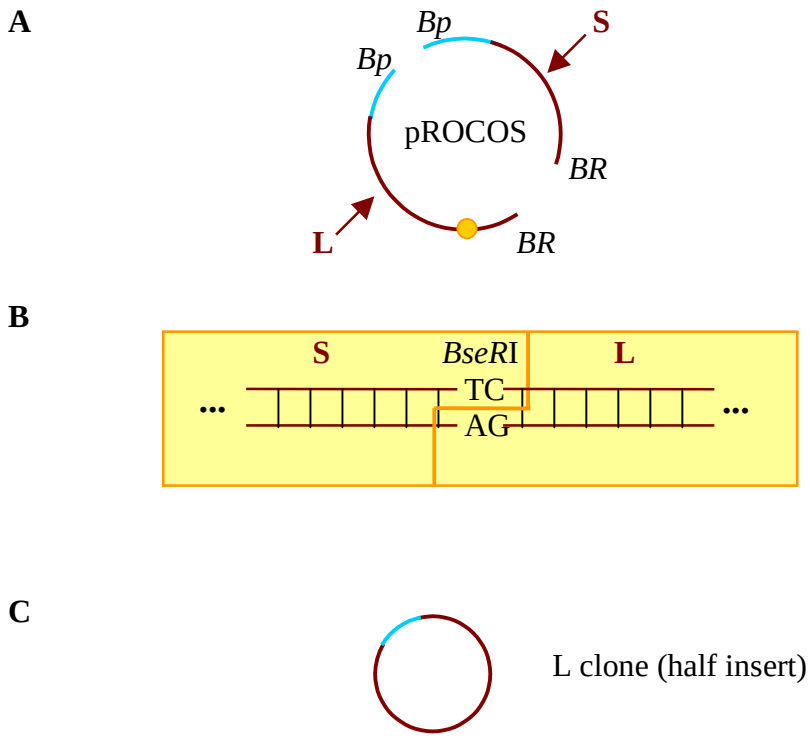


Figure III-18: Formation of the unproductive L clone. (A-) pROCOS vector after restriction with *Bpm*I and *Bse*RI, which form the S and L fragments. (B-) details of the *Bse*RI restriction site within pROCOS. (C-) formation of the L clones, by reclosure of the F fragment on its own, when the TC set was used on the *Bpm*I restriction site. L clones miss a part of the vector (approximately half of the pIII) and contain only half of the insert.

The L clones were composed of incomplete vectors. They lacked the S fragment and, therefore, around half of the pIII gene was missing, preventing peptide display on the phage particles. These L clones, however, have one advantage. Since they are smaller than the usual vectors of the library, they have a selective advantage due to a shorter replication time. Thus, they are an excellent indicator of the quality or of the chance of success of a panning selection. Indeed, if no ligand is isolated against a particular target, the L clones should

become enriched, due to their selective advantage and should after a few cycles of panning dominate over other clones. Hence, the number of L clones present in successive panning rounds, is an indication of the progress of the affinity selection.

Within the 91% of clones appropriate for panning (containing the insert, or the insert plus the proline rich cassette), nearly 85% are correctly formed (27 clones; Table III-11). The other 15% were unproductive clones due to changes in the reading frame (insertions, or deletions) or to the presence of a stop codon.

	number of clones	**percentage**
Cassette/insert	25	71,4
Insert	7	20
Cassette/insert + insert	**32** (25+7)	**91,4** (71,4+20)
L clones	3	8,5
Correct clones	**27**	**84,3**
Insertions	2	6,2
Deletions	1	3,1
Stop codons	2	6,2

Table III-11: DNA sequence analysis of 35 randomly selected clones from CPLPPXP library. Relative frequency of correctly and incorrectly formed clones. The first part of the table relates to the proportion of clones which present the cassette within the insert, only the insert, or which are malformed. The bottom part of the table presents a further analysis of the 32 clones correctly formed, to check the DNA quality: insertion, deletions, or presence of stop codon.

The 27 clones containing the hypervariable insert, with or without proline-rich cassette, which show a correct reading frame and no stop codon, were further investigated to determine the amino acid frequency. The results of the analysis are presented in Table III-12, where they are compared with the theoretical frequency expected in a NNB library. The Figure III-19 gives an overview of the data presented in Table III-12.

The amino acid frequency estimated for the cosmix-plexing® CPLPPXP library is very similar to the expected values of a NNB bank, as seen in Table III-12. The only unusual data are those obtained for Trp and Phe which were both over-represented, and for Thr and Ser, which were under-represented. The frequency of different amino acids in the CPLPPXP library, are nevertheless convenient to carry on panning selection, against the SH3 and EVH1 domain targets.

	NNB	CPLPPXP
Ala	6	5,6
Arg	8	8,1
Asn	4	4
Asp	4	3,4
Cys	4	5,3
Glu	2	1,8
Gln	2	3,4
Gly	6	6,8
His	4	2,8
Ile	4	5,9
Leu	8	10,3
Lys	2	1,8
Met	2	1,8
Phe	4	8,4
Pro	6	7,5
Ser	10	5,6
Thr	6	3,1
Trp	2	5
Tyr	4	3,4
Val	6	4
Stop codon	2	1,2

Table III-12: Expected and actual amino acid frequencies (in percentage) of the CPLPPXP library. The NNB values correspond to the theoretical amino acid frequency expected for libraries constructed with the codon scheme NNB, where N is A, C, G, or T, whereas B stands for C, G, or T.

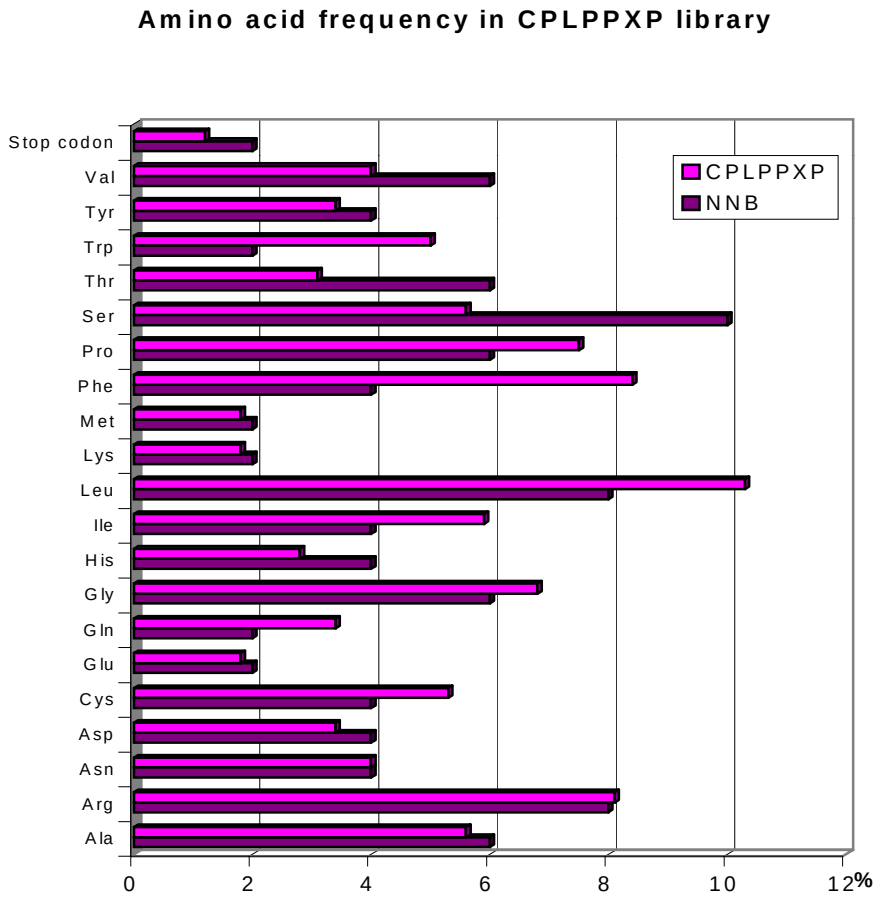


Figure III-19: Expected and actual amino acid frequencies (in percentage) of the CPLPPXP library, according to the values presented in Table III-12.

D - Panning selection on SH3 domains

The CPLPPXP library is composed of an enormous collection of phage variants ($2x10^8$) which display, N-terminally on the pIII protein, linear biased random peptides. The amino acid sequence of the exposed random peptides has been conceived to interact favorably with SH3 and EVH1 domains: Xxx_5 Pro Pro Xxx Pro Xxx_4. A major characteristic of this library is that it has been designed as a cosmix-plexing® library. This means that at any time, it is possible to recombine the left and the right parts of the peptides, in order to increase the affinity of the ligands for their targets.

The quality of the CPLPPXP bank was considered adequate. The diversity of the bank is slightly under the theoretical size required, but this drawback is largely compensated by the cosmix-plexing® recombination possibilities. The library contains codons for all amino acids (even if a few of them are over- or under-represented), except two of the stop codons. Therefore, the bank is relevant to be used in the selection of ligands binding to SH3 and EVH1 domain targets.

Before extending the study to the SH3 and EVH1 domains, the optimal panning parameters had to be determined, and therefore, the effort was first focused on only one target, the Src SH3 domain.

1- Src SH3 domain: a validation target

Src was selected as the model of choice to characterize the best conditions for the experiments, since it has been previously extensively studied. Thus, the idea of the type of ligands expected to be isolated was quite precise.

The CPLPPXP library used in this work to isolate proline rich ligands against the Src SH3 domain, looks very similar to some of the banks previously used by other researchers. However, although it was expected to characterize ligands showing high consensus sequence homologies, compared to the Src SH3 ligands previously described, it was hoped to find ligands which show higher binding affinity for their target.

1-1 Optimization of the affinity selection procedures

Extended tests were performed on Src SH3 domain, to determine the panning selection conditions which lead to the best results.

The panning procedure was performed as described for the constrained libraries, however, some steps were repeated with modification, in order to encourage the interaction between the target protein, and the phage library. These steps are emphasized here.

The immobilization of the Src SH3 domain was performed at 4°C overnight, in PBS, but two kinds of supports were compared: a maxisorb (Nunc), and a glutathione microtiter plate (Pierce). On the glutathione plates, the GST is captured by the glutathione fixed to the bottom of the wells, and the fused protein (Src) is expected to be orientated such that it is available for interactions with the phage library.

A

	cycle of panning		
	1	2	3 - 5
PBS/Tw 0,05	1x	3x	5x
PBS/milk	10 min	10 min	10 min
PBS/Tw 0,05	1x	3x	5x
PBS/milk	5 min	5 min	5 min
PBS/Tw 0,05	1x	3x	5x
H_2O	1x	1x	1x

B

	cycle of panning		
	1	2	3 - 5
PBS/Tw 0,05	3x	5x	5x
PBS/milk	10 min	10 min	10 min
PBS/Tw 0,05	3x	5x	5x
PBS/milk	5 min	5 min	5 min
PBS/Tw 0,05	3x	5x	5x
PBS/Tw 0,05			5x
H_2O	1x	1x	1x

C

	cycle of panning		
	1	2	3 - 5
PBS/Tw 0,5	1x	3x	5x
PBS/milk	10 min	10 min	10 min
PBS/Tw 0,5	1x	3x	5x
PBS/milk	5 min	5 min	5 min
PBS/Tw 0,5	1x	3x	5x
H_2O	1x	1x	1x

Table III-13: Stringency of the washing steps according to the cycle of panning. (A-) Basic procedure for the washing step. The PBS/Tween contains 0,05% of Tween 20. The PBS/milk, is the blocking solution, with 2% of skim milk powder. Incubation with the blocking solution tends to reduce the background level. It can happen that phage bind to the milk powder used to block the walls of the wells, the milk added in solution competes this interaction. (B-) On this procedure, the stringency of the washing is increased, by washing several time more. (C-) This protocol is also based on higher stringency, but by using a higher concentration of Tween (0,5%), instead of multiplying the number of times the washing are performed. Each time the washing was achieved, the incubation was prolonged for 5 min. Tw stands for Tween 20.

After immobilization, the wells were blocked with PBS/2% skim milk powder, to prevent unspecific binding of the phage to the walls of the wells, to keep background levels low. CPLPPXP library was incubated in the wells, after the blocking. Previously, the phage were diluted 1 to 1 in PBS/2% skim milk powder, to reduce the background. Indeed, phage which show affinity for milk, would interact in solution, and would be, later, washed away more easily.

After incubation between the target and the phage library, the wells were washed, to remove non binding, or low affinity phage. The stringency of the washing steps increase with the cycles of selection. The washing step is a crucial stage of the panning, since it is the one where it is possible to remove the background. However, an equilibrium has to be found, in order to avoid washing away clones which might be able to bind the target. Therefore several washing conditions were tested (Table III-13), to determine which is adequate for the CPLPPXP library, and the Src SH3 target. Three washing procedures were tested, with different stringency.

Each of the three types of washing (A, B, C; Table III-13) were tested on both glutathione and maxisorb microtiter plates, when CPLPPXP library was screened against the Src SH3 domain. As a control, CPLPPXP library was incubated in an empty well, where PBS alone was coated. This control was performed in both plates, and the washing procedure used was from the type A (Table III-13), the less stringent washing conditions used in this series of tests. The progression of the output values (eluted phage at the end of the panning procedure) can be followed in the Figure III-20.

The comparison of the output values presented in Figure III-20, gives an impression of the enrichment during the panning. The glutathione and maxisorb microtiter plates show very different evolutions, and surprisingly, the maxisorb plate seems to give more promising results.

After the first round of panning selection, more phage were recovered from the glutathione plate (same amount of phage incubated with Src, everywhere). The dramatic decrease after the second round of panning, observed with all washing conditions, in both plates, is explained by the loss of the background. After the first round, the majority of the non binder clones have been rejected. However, the one still present, and eluted at the end of the affinity selection, were amplified, before being incubated with Src for the second round. The observed decline is thus an indicator that this amplified background have been washed away.

A

	Glutathione plate			
	A	B	C	Blank
1R	5,00E+06	3,00E+05	1,30E+06	6,00E+06
2R	1,20E+03	2,10E+03	5,00E+03	1,30E+03
3R	1,40E+04	2,00E+02	7,00E+02	1,30E+04
4R	1,10E+04	3,00E+03	2,20E+03	4,60E+03

B

	Maxisorb plate			
	A	B	C	Blank
1R	1,30E+04	2,00E+04	2,70E+04	1,00E+05
2R	1,20E+03	1,60E+03	4,00E+03	2,50E+04
3R	2,50E+05	7,00E+03	1,80E+04	4,00E+04
4R	2,20E+06	8,00E+04	8,00E+05	3,40E+04

C

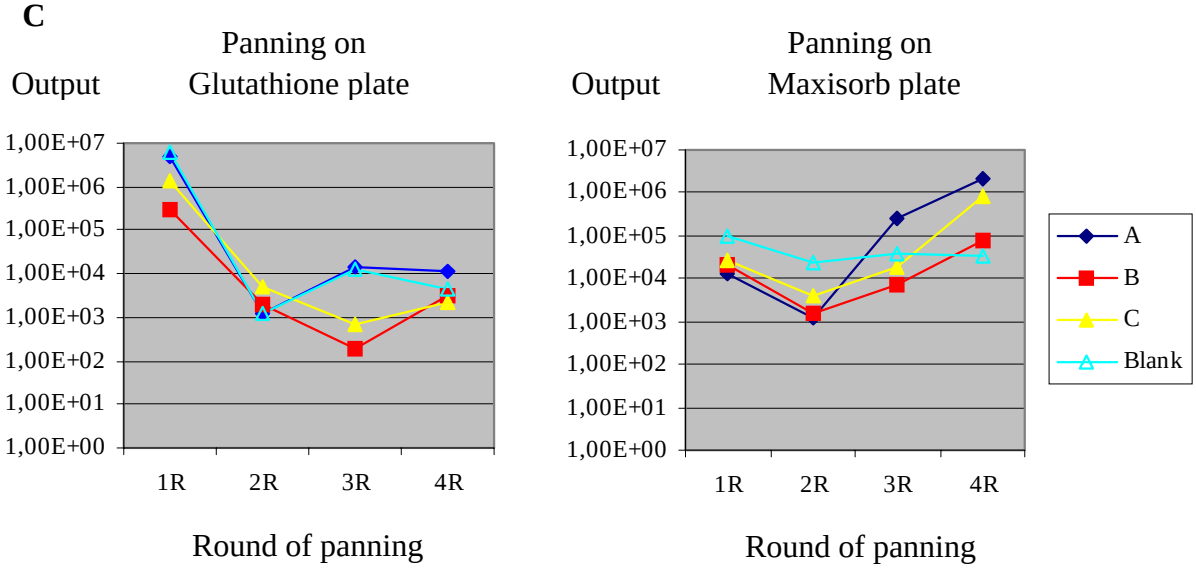


Figure III-20: Evolution of the output titer during the screening of CPLPPXP on Src SH3 domain. (A and B-) Progression of the output values, in absolute number, after 4 rounds of panning (noted 1R to 4R). The same amount of phage were incubated with Src, in all tests ($2x10^{11}$). The A, B, and C, refer to the washing procedures used, as they are described in Table III-13. The blank is an empty well where the CPLPPXP library was incubated, and then, washed using the protocol A. The tables A and B described the output values of the screening on glutathione or maxisorb microtiter plates, respectively. (C-) Diagram of the values observed in both A and B tables.

From the third round on, the maxisorb and glutathione plate outputs, diverge in their profile. In the glutathione plate, the wells washed under the B and C conditions, carry on decreasing in the output number, and slightly recovered on the fourth round, while the well washed under the A procedure (less stringent), recovered already after the third round, but then, get stabilized on the fourth cycle. However, with all three kinds of washing, the output values after the fourth round, stay much under the values from the first round. In addition,

they remain very similar to the blank output value. This might indicate that no selection occurs during the four cycles of panning. The output values of the library incubated with Src, and washed under the A conditions, follow nearly perfectly the values of the control (which was washed the same way). This suggests that the GST-Src protein was not retained on the glutathione of the microtiter plate, or that the interaction between the glutathione and the GST, somehow masked the Src, which would not be available anymore for interactions with the CPLPPXP library.

		Input	Output	Output/Input	Enrichment
A	1R	2,00E+11	1,30E+04	6,50E-08	-
	2R	1,00E+10	1,20E+03	1,20E-07	-
	3R	3,20E+10	2,50E+05	7,80E-06	120
	4R	2,20E+10	2,20E+06	1,00E-04	1538
B	1R	2,00E+11	1,30E+04	6,50E-08	-
	2R	6,00E+09	1,20E+03	2,00E-07	-
	3R	5,00E+10	2,50E+05	5,00E-06	77
	4R	9,00E+10	2,20E+06	2,40E-05	369
C	1R	2,00E+11	2,70E+04	1,30E-07	-
	2R	1,30E+10	4,00E+03	3,00E-07	-
	3R	3,20E+09	1,80E+04	5,60E-06	43
	4R	7,00E+10	8,00E+05	1,10E-05	84

Table III-14: Enrichment factors of CPLPPXP screened on Src SH3 domain, on maxisorb plate, according to the type of washing conditions used (A, B, or C, described in Table III-13). The evolution of the input (amount of phage incubated with the target protein), output (eluted phage after stringent washing steps), and the ratio output divided by input (O/I) are shown, along the cycles of panning. The O/I ratio is a parameter indicating if an enrichment occurred. When most of the phage given as input are recovered after washing step and elution (output), that means that the largest majority of the input presented affinity for the target, and remained bound to the target during the washing steps. These clones have been enriched during the previous cycles of panning selection. The enrichment factor is estimated by dividing the ratio O/I from the round of interest (where the output and the ratio O/I appear to be very high), with the ratio O/I from the first round. 1R, 2R, 3R and 4R, stand for first, second, third, and fourth rounds of panning.

As far as the maxisorb microtiter plate is concerned, from the third round on, the output values significantly increased, revealing an enrichment, since the values are higher than they were after the first round, and in addition, above the control values (after the fourth cycle). The enrichment factors, which take into account the input (amount of phage incubated with the target protein), are precisely calculated, and presented in Table III-14. The washing condition which appears to be the most favorable for the selection, is the less stringent one (A,

described in Table III-13). Indeed the output values of panning washed under the A procedure, get higher than the one of the first round, and of the control, already after the third round.

1-2 Isolated ligands against Src SH3 domain

To confirm the hypothesis about the quality of the affinity selection, according to the panning procedure used, single clones from the third and fourth cycles, were randomly selected, and sequenced.

Whatever the washing conditions, and the round of panning, the clones selected from the glutathione plate show no enrichment. 40% of the sequenced clones were L clones (vector only partially regenerated, as shown in Figure III-18). 25% contained stop codons, 15% had a modification in the reading frame, due to 1 bp insertion, or deletion. Finally, only 20% of the clones show a normal displayed peptide, but without any homology among them.

The maxisorb clones confirmed the enrichment observed after the third and fourth rounds. 10 clones from each round, and each washing protocol were sequenced. The resulting sequences are summarized here. Some clones are unproductive (when insertion, deletion, or stop codons occur, or when a L clone is found), however, most of them are correctly formed, and are classified in the Figure III-21, regarding their occurrence.

3R

A *1x clone with 1 bp deletion*
2x L clones

6x S2	Arg	Pro	Leu	Pro	Leu	Pro	Pro	Val	Pro	Trp	Val	Arg	Trp
1x S5	Ser	His	Tyr	Ser	Leu	Pro	Pro	Gly	Pro	Leu	Arg	Gly	Phe

B *3x L clones*
2x clones with stop codon
1x clone with 1 bp insertion

4x S2	Arg	Pro	Leu	Pro	Leu	Pro	Pro	Val	Pro	Trp	Val	Arg	Trp

C

9x S3	Arg	Ser	Leu	Pro	Leu	Pro	Pro	Val	Pro	Thr	Ser	Ser	Ala
1x S4	Lys	Gly	Arg	Pro	Leu	Pro	Pro	Val	Pro	Gly	Thr	Pro	Ser

4R

A

8x S1	Gly	Asn	Arg	Pro	Leu	Pro	Pro	Ile	Pro	Ser	His	Pro	Phe
1x S7	Gly	Ala	Gly	Ser	Leu	Pro	Pro	Cys	Pro	Ala	Asp	Leu	Leu
1x S6	Asn	Ala	Cys	Thr	Leu	Pro	Pro	Phe	Pro	Trp	Asp	Arg	Cys

B *2x L clones*
1x clone with 1 bp deletion
1x clone with stop codon

6x S2	Arg	Pro	Leu	Pro	Leu	Pro	Pro	Val	Pro	Trp	Val	Arg	Trp

C *3x clones with 1 bp deletion*
2x L clones

5x S3	Arg	Ser	Leu	Pro	Leu	Pro	Pro	Val	Pro	Thr	Ser	Ser	Ala

The alignment of the isolated and sequenced clones (Figure III-21), obviously shows a consensus sequence Arg Pro Leu Pro Pro Ψ Pro Ψ (where Ψ represents hydrophobic amino acids), similar to the one expected for class I ligands, with respect to the work accomplished by other laboratories. Class I ligands are assumed to be the dominant class of ligands for Src SH3 domain.

	Sequences against Src SH3 domain														Frequency (3R	4R)
S1	Gly	Asn	**Arg**	**Pro**	Leu	**Pro**	**Pro**	**Ile**	**Pro**	Ser	His	Pro	Phe		-	8
S2			**Arg**	**Pro**	Leu	**Pro**	**Leu**	**Pro**	**Pro**	**Val**	Pro	Trp	Val	Arg Trp	10	6
S3			**Arg**	**Ser**	Leu	**Pro**	**Leu**	**Pro**	**Pro**	**Val**	Pro	Thr	Ser	Ser Ala	9	5
S4	Lys	Gly	**Arg**	**Pro**	Leu	**Pro**	**Pro**	**Val**	**Pro**	Gly	Thr	Pro	Ser		1	-
S5	Ser	His	Tyr	**Ser**	Leu	**Pro**	**Pro**	Gly	**Pro**	**Leu**	Arg	Gly	Phe		1	-
S6	Asn	Ala	Cys	Thr	Leu	**Pro**	**Pro**	**Phe**	**Pro**	**Trp**	Asp	Arg	Cys		-	1
S7	Gly	Ala	Gly	**Ser**	Leu	**Pro**	**Pro**	Cys	**Pro**	**Ala**	Asp	Leu	Leu		-	1
Consensus sequence																
	Xxx	Xxx	**Arg**	**Pro/S**	Leu	**Pro**	**Pro**	**Ψ**	**Pro**	**Ψ**	Xxx	Xxx	Xxx			

Figure III-21: Sequences of single clones, after 3rd and 4th rounds of panning against Src SH3 domain, and the deduced consensus sequence. The clones are classified according to their frequency. The first Pro of the consensus sequence appears, in the single clones, as often as a Ser. However, the Pro occurs mainly in the strongest enriched clones, demonstrating its significance. In addition, in term of flexibility, Ser presents similarity to Pro. A residue is considered as representative for the consensus sequence, when it occurs in 50% or more of the sequences. Underlined amino acids are fixed residues within the CPLPPXP library. Ψ represents hydrophobic amino acids.

The clones which present the most homology to the conceived consensus sequence, are the one which occur the most often, confirming the good efficiency of the enrichment. The clones that appears only once, might be present only by chance.

The first determined Pro of the consensus sequence appears, in the single clones, as often as a Ser residue. However, the Pro occurs mainly in the strongest enriched clones, demonstrating its higher significance. In addition, in term of flexibility, Ser presents similarity to Pro.

The S5 clone shows characteristics belonging to the class II ligands: the position of the Arg after the proline rich motif. However, this clone does not completely correspond to the

consensus sequence conceived for class II ligands (PPLPXR), since the position between the prolines is not a leucine, and moreover, is not a hydrophobic residue. This clone is not expected to show strong reaction with Src, since it occurs only once in the third round, and disappears in the fourth one.

The conception of a consensus sequence similar to the one defined by other groups, confirmed that the CPLPPXP library is appropriate to isolate and characterize binders against the SH3 domains, and that the strategy applied for the panning selection is adequate**.**

It is interesting to note that the three best clones (S1, S2, and S3) were isolated from different panning protocols. This demonstrates that using different washing methods, more or less stringent, lead to the obtention of diverse ligands. Therefore it might be advantageous to analyze the effect of different washing conditions, each time a new protein serves as target in an affinity selection. However, to facilitate the work, one protocol has been selected, to be used for the primary panning of all targets.

The maxisorb microtiter plate is the material selected for the next panning procedure, since, the glutathione plate did not allow discrimination between non-, weak-, or strong-binders.

The three washing conditions tested, yield good results, when combined with the maxisorb plate. However, in future work, the procedure A will be preferred, since it leads to the most favorable enrichment (at least for the Src SH3 domain, Table III-14 and Figure III-21).

Therefore, the panning procedures performed on different SH3 and EVH1 domains were achieved on maxisorb microtiter plates, and washed mildly, with PBS containing 0,05% Tween 20, and with a reduced number of washing cycles (as described under the A conditions, in Table III-13).

1-3 Cosmix-plexing® recombination

During the first panning round, the future dominant clones (S1, S2, S3) were present in a small amount, as well as many other clones which exhibit weak affinity properties for the target. The dominant clones get enriched, during further rounds, and compete with the weaker binders which are then eliminated during the washing steps of the panning.

The preselected populations eluted after the first and second rounds of panning, are the working material submitted to the recombination. At this stage, most of the clones present are

able to show an affinity for the target, even if very weak. The hypervariable region of the clones in this preselected population are recombined together, producing all possible recombinants. By reassorting these clones among themselves, one links together two parts of two different clones (Figure I-6, I-7). This raises the chance to bring together initially weak domains, but which together exhibit new features which may correspond to the "perfect" structure for the binding to the target (synergism). Further selection with this optimized library is expected to lead to the isolation of new clones which should be optimal regarding their affinity to the target.

Cosmix-plexing® is performed from maxiprep prepared from a glycerol stock aliquot of resuspended cells infected with eluted phage, after the first and second rounds of panning. The two subpopulations were mixed together in a ratio one to one (i.e., 50% from the clones came from the first round, and 50% from the second round).

The strategy of this approach is described in Figure III-22. The DNA of the preselected population, recovered from the maxiprep as plasmids, is restricted by *Bpm*I, opening the hypervariable region is the middle. After purification, the opened vectors are ligated at high DNA concentration (>200 µg/ml) leading to the formation of concatemers. This ligation ensures a recombination among the vectors, increasing the diversity of the hypervariable region. The ZZ set used at the opened position of the variable domain, does not allow head to head joining of fragments.

The concatemers are resolved into single phagemid molecules on cleavage with *Bgl*I, followed by ligation at a low DNA concentration (<40µg/ml). The optimized library was electroporated, and packaged.

The packaged cosmix-plexed library was screened against Src SH3 domain on affinity selection. Randomly selected clones were sequenced after the first round of panning, as our experience demonstrated that after cosmix-plexing®, enriched clones are efficiently obtained from the first or second cycles of selection.

The 10 clones which were sequenced, were all identical. The enriched variant was the S1 sequence, previously enriched on panning selection, indicating a very strong binding preferences of Src SH3 domain for this ligand.

S1 Gly Asn **Arg Pro Leu Pro Pro Ile Pro** Ser His Pro Phe

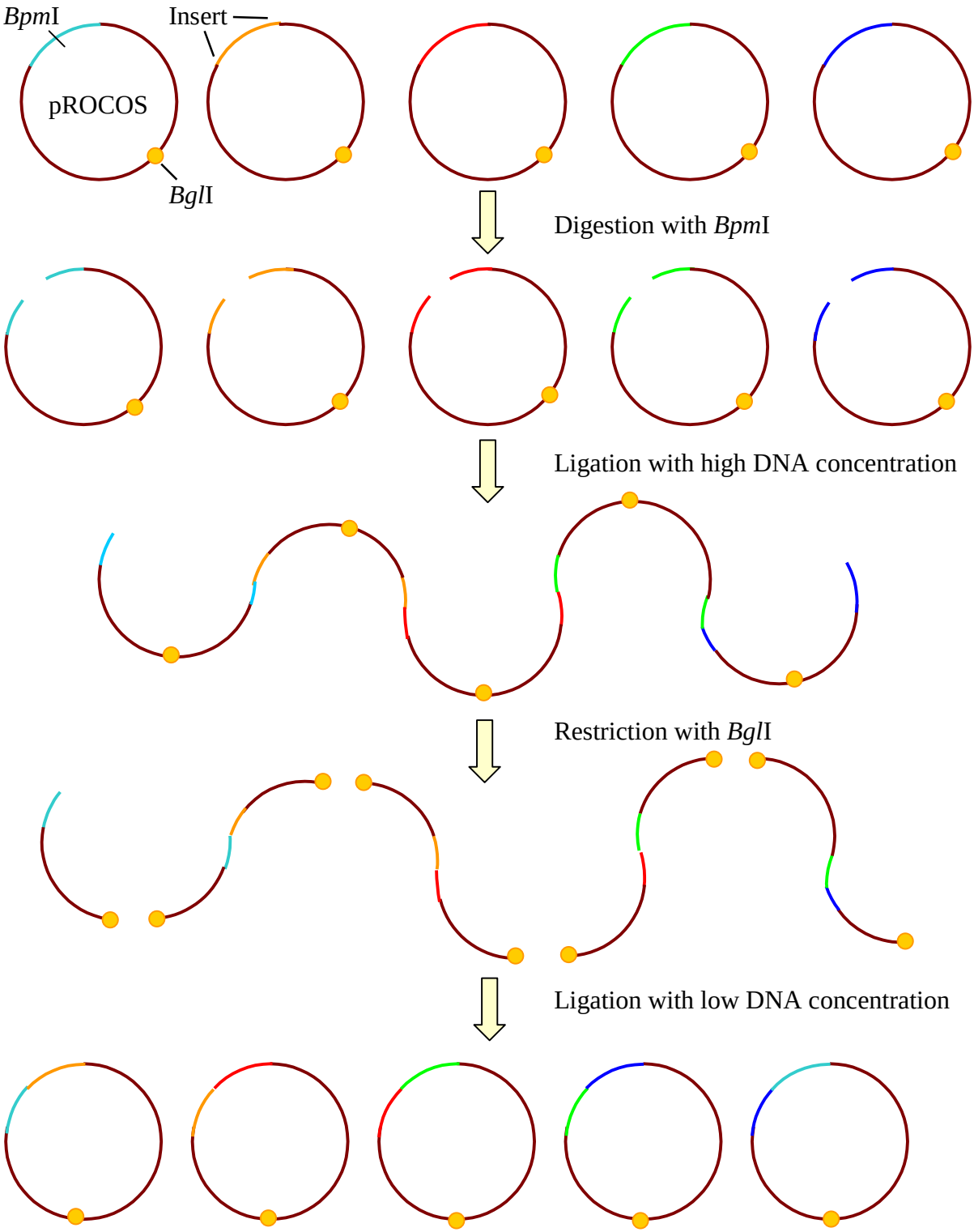


Figure III-22: Schematic representation of cosmix-plexing® recombination. The initial material, is phagemids obtained by maxiprep of a glycerol stock aliquot, which were prepared from the resuspended colonies infected with the eluted phage from the first and second rounds of panning. The vectors are opened with *Bpm*I, and religated at high DNA concentration to yield to the formation of concatemers. The concatemers are resolved as single vectors by restriction with *Bgl*I, and ligated with low DNA concentration, as closed monomers. The recombined vectors can be electroporated, packaged, and used for further panning to isolate optimized ligands against a particular target.

1-4 Phage ELISA

The clones characterized after the panning (Figure III-21), were propagated as single phage, and individually checked on phage ELISA for affinity against the Src SH3 domain. The wells of an ELISA microtiter plate were coated with the GST-Src protein. A well, coated with GST alone, was used as control to ensure that the single phage interact with the Src, and not with the GST (not shown). The results are summarized in Figure III-23.

On phage ELISA, only four of the clones react positively with the Src domain, S1 to S4. These clones are the one which contain the **Arg Pro(/Ser) Leu** sequence, before the proline motif, demonstrating the crucial role of this sequence for the binding of the ligands to the Src SH3 domain.

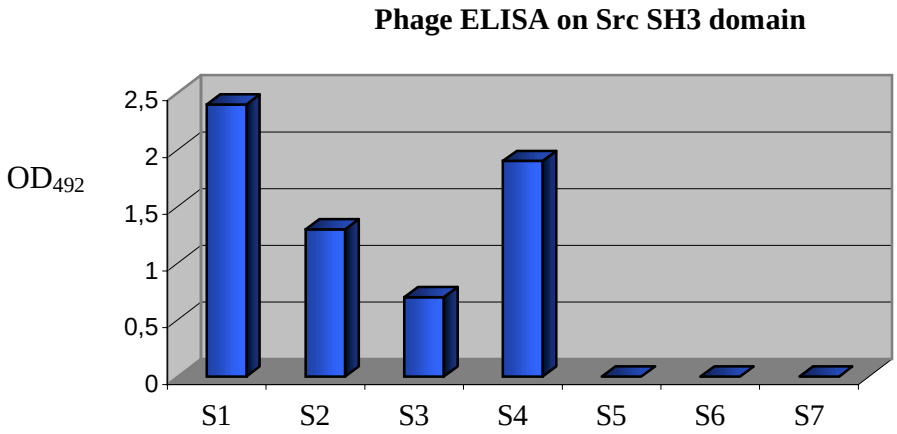
As expected, the S5 clone, which was assimilated to a class II ligand does not bind to the Src domain, along with the S6 and S7 peptides. Since they could not get enriched in the fourth panning, it was obvious that the dominant clones (S1, S2, S3) would have bind better to the Src domain.

More surprising is the clone S4, which appeared once in the third round, but not any more in the fourth round. Since S1, S2 and S3 clones dominate over the other ones, in the fourth round, it was expected that S4 binds weakly, or not at all to the Src SH3 domain, similarly to the clones S5, S6 and S7 which also have a low occurrence. However, the S4 clone is one of the best Src ligand, according to the relative binding of the peptides.

Among the four clones which interact with Src, it is interesting to see that the weaker one, the S3, is the only one of the four which does not exhibit Arg Pro Leu before the proline motif, but rather, Arg Ser Leu. This is a very important clue to testify the significant influence that the Pro have over the Ser, on the binding affinity of a ligand for the Src domain.

The alignment of the four positive clones allows the conception of a new consensus sequence, which might be more exact, since the clones aligned were selected as the one which were proved to react with Src. This consensus sequence is thus more precise, and clearer than the one established after the panning. The consensus is also more extended. The new consensus sequence corresponds to the preferred properties required by a peptide, to be able to bind strongly to the Src domain, and shows the following sequence (where Ψ represents the hydrophobic residues, and the underlined amino acids were previously fixed in the library):

Arg Pro Leu Pro (Pro/Leu) Ψ Pro Val Pro Pro S/Ψ



	Phage ELISA against Src SH3 domain (positive clones)														OD492
S1	Gly	Asn	**Arg**	**Pro**	**Leu**	**Pro**	**Pro**	**Ile**	**Pro**	Ser	His	**Pro**	**Phe**		2,4
S4	Lys	Gly	**Arg**	**Pro**	**Leu**	**Pro**	**Pro**	**Val**	**Pro**	Gly	Thr	**Pro**	**Ser**		1,9
S2			**Arg**	**Pro**	**Leu**	**Pro**	**Leu**	**Pro**	**Pro**	**Val**	**Pro**	Trp	**Val**	Arg Trp	1,3
S3			**Arg**	**Ser**	**Leu**	**Pro**	**Leu**	**Pro**	**Pro**	**Val**	**Pro**	Thr	**Ser**	Ser Ala	0,7
Consensus sequence															
	Xxx	Xxx	**Arg**	**Pro**	**Leu**	**Pro**	**P/L**	**Ψ**	**Pro**	**Val**	**Pro**	**Pro**	**S/Ψ**		

Figure III-23: Relative affinity, according to a phage ELISA test, of the clones isolated against Src SH3 domain. GST-Src protein was coated in the wells. An equivalent amount of phage were incubated in every well (10^{10} phage/well). After several washes, the remained bound phage were detected with anti-M13 antibody. The absorbance was detected at 492 nm. A well coated with GST alone was used as a control, to ensure that the single clones bind to the Src, and not to the GST. The phage showed no interaction with the GST (not shown). The sequence of the clones which react positively with the Src domain are shown, and a new consensus sequence is conceived. The OD values are an average of three repeated tests. Underlined residues are fixed within the CPLPPXP library. Ψ represents the hydrophobic residues.

Since the clone S1 was the most enriched after the fourth round, and since it was the only one present after one round of affinity selection after cosmix-plexing®, it was expected, to interact the most strongly with the Src domain. The phage ELISA demonstrates that it is exact. Therefore, this clone was selected for further analysis. S1 was investigated to explore further the capacity of SH3 domains to discriminate between SH3 ligands. Therefore, S1

clone was studied to search for cross reaction with a panel of SH3 domains. It was also interesting to check the potential ability of the clone S1, to bind to EVH1 domains, since these domains are also known to interact with proline rich ligands.

Equal titers of the clone S1 were incubated in microtiter wells coated with different GST-SH3 fusion proteins. After several washes, the bound phage were detected with an anti-M13 antibody (Figure III-24).

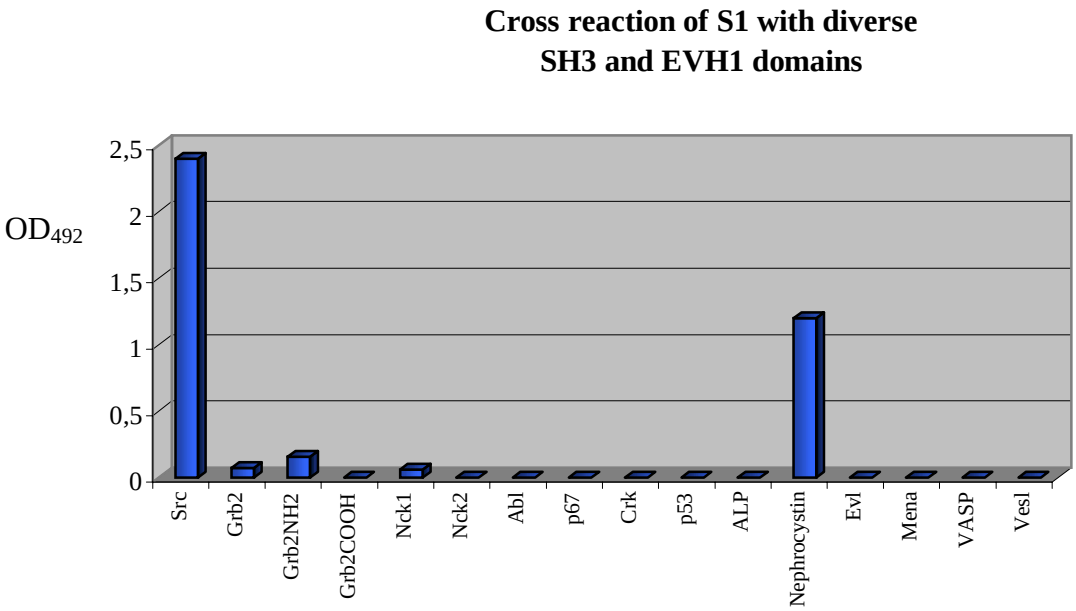


Figure III-24: Cross reaction of the clone S1 for a panel of SH3 and EVH1 domains, tested on phage ELISA. Cross reaction with EVH1 domains (Evl, Mena, VASP, and Vesl) were also examined, since the EVH1 domains are also known to bind to proline rich ligands. All SH3 and EVH1 domains were coated as GST fusion proteins. 10^{10} phage of the clone S1 were incubated in the wells. After several washes, the remained bound phage were detected with anti-M13 antibody. The absorbance was detected at 492 nm.

The clone S1 shows equivalent ELISA signal for Src, to those obtained previously. No significant cross reaction is detected, except with the Nephrocystin SH3 domain, but this will be studied in more details later, along with the clones isolated against the Nephrocystin protein, since the consensus sequences of ligands which bind Src or Nephrocystin are very similar. S1 reveals an excellent specificity for the Src and the Nephrocystin SH3 domain.

None of the four EVH1 domains immobilized on the ELISA plate interact with the S1 clone.

2- Nck2 SH3 domain

The Nck protein has been kindly provided by Lawrence Quilliam and Brian Kay (University of Wisconsin-Madison, Wisconsin, USA), as GST constructs, which once expressed, represent the three SH3 domains separately (GST-Nck1, GST-Nck2, GST-Nck3). Unfortunately, for a unknown reason, the expression of the GST-Nck3 construct resulted in the expression of the GST alone.

2-1 Panning procedure

GST-Nck1, and -Nck2, were correctly formed, and highly purified due to the GST sepharose columns (as previously described). Both of the domains were immobilized on maxisorb microtiter plate, and incubated with the CPLPPXP library. The panning procedure was performed as developed for Src SH3 protein, with mild washing steps, but a little bit more stringent every new cycles.

No significant enrichment was observed. However, the sequencing of some randomly selected clones, after the fifth round of panning, leads to a short consensus sequence. Even if approximately half of the clones show some changes in their reading frame, or stop codons, two clones occur several times, for both Nck1 and Nck2. Curiously, these two clones belong to the 20% of the library which does not contain the proline rich cassette. Therefore, these clones were called Non Proline 1 and 2 clones (NP1 and NP2):

NP1 Ser Thr Leu **Cys Asp Gly Tyr Cys**

NP2 Asn Pro Met Gly **Asp Gly Tyr** Phe

These 8 amino acids long peptides show a 3 residue stretch consensus: **Asp Gly Tyr**, which might then be favorable for the binding to Nck1 and Nck2 SH3 domains. It is however unclear if these clones are really significant, since SH3 ligands are expected to form a polyproline type II helix, and to contain a proline rich motif (Pro Pro Xxx Pro motif) to be able to interact with the SH3 domain.

2-2 Cosmix-plexing® recombination

A cosmix-plexing® recombination was carried out, to better analyze the binding preferences of the Nck domains. The populations recovered after the first and the second rounds of the primary panning, for both domains, were submitted to cosmix-plexing®, as previously described (Figure III-22).

Sequences against Nck2 SH3 domain, from cosmix plexing library

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}	P_{13}
First group													
	Ser	Leu	Gly	**Ile**	Trp	**Pro**	**Pro**	**Phe**	**Pro**	**Trp**	Ile	**Pro**	**Ile**
	Ala	Ser	**Cys**	**Phe**	Gln	**Pro**	**Pro**	**Phe**	**Pro**	**Trp**	Thr	**Pro**	**Leu**
	Arg	His	**Cys**	**Val**	Gly	**Pro**	**Pro**	Asp	**Pro**	**Tyr**	Val	**Pro**	**Met**
	Thr	Cys	**Thr**	**Ala**	Leu	**Pro**	**Pro**	**Phe**	**Pro**	**Trp**	Ser	**Pro**	Arg
	Leu	Val	**Thr**	Cys	Val	**Pro**	**Pro**	**Phe**	**Pro**	**Trp**	Pro	**Pro**	**Pro**
Consensus sequence													
	Xxx	Xxx	**C/T**	**Ψ**	Xxx	**Pro**	**Pro**	**Phe**	**Pro**	**W/Y**	Xxx	**Pro**	**Ψ**
Second group													
	Pro	Arg	Asn	Val	**Arg**	**Pro**	**Pro**	Ala	**Pro**	**Trp**	His	Gly	**Ser**
	Ser	Thr	Gly	**Arg**	**Arg**	**Pro**	**Pro**	His	**Pro**	**Phe**	**Leu**	Val	**Ser**
	Lys	Phe	Met	Asn	**Arg**	**Pro**	**Pro**	Gly	**Pro**	**Trp**	**Leu**	Ala	**Cys**
	Ala	Ser	Ser	Thr	**Arg**	**Pro**	**Pro**	Cys	**Pro**	**Tyr**	**Leu**	Asp	Ile
	Arg	Pro	Arg	**His**	**Arg**	**Pro**	**Pro**	Val	**Pro**	Val	**Val**	Lys	Pro
	Ser	Val	Phe	Leu	**Arg**	**Pro**	**Pro**	Cys	**Pro**	Ala	**Leu**	Gly	Asn
	Ser	Ser	Tyr	**Arg**	**Arg**	**Pro**	**Pro**	Leu	**Pro**	Ser	His	Arg	**Cys**
Consensus sequence													
	(Ser)	Xxx	Xxx	**(β)**	**Arg**	**Pro**	**Pro**	Xxx	**Pro**	**φ**	**Leu**	Xxx	**S/C**
Third group													
	Met	Phe	Gln	**Phe**	Gly	**Pro**	**Pro**	Cys	**Pro**	**Tyr**	**Leu**	Asp	**Ile**
	Val	Gly	**Leu**	His	Trp	**Pro**	**Pro**	**Val**	**Pro**	**Trp**	**Val**	Gln	Arg
	Leu	Lys	**Leu**	**Val**	Asn	**Pro**	**Pro**	Gln	**Pro**	**Tyr**	**Phe**	**Ile**	**Leu**
	Asp	Ser	**Leu**	**Ala**	Glu	**Pro**	**Pro**	**Phe**	**Pro**	**Tyr**	**Met**	**Ile**	**Val**
	Gly	Glu	Tyr	**Val**	Trp	**Pro**	**Pro**	**Val**	**Pro**	**Trp**	Ser	**Met**	**Met**
Consensus sequence													
	Ψ	Xxx	**Leu**	**Ψ**	Xxx	**Pro**	**Pro**	**Ψ**	**Pro**	**W/Y**	**Ψ**	**Ψ**	**Ψ**

Figure III-25: Sequences of single clones, after two rounds of cosmix-plexing® panning against Nck2 SH3 domain, and the deduced consensus sequences. The clones occurred only once each. A residue is considered as representative for the consensus sequence, when it occurs in 50% or more of the sequences. Underlined amino acids are fixed residues within the CPLPPXP library. Ψ, β, and φ represent hydrophobic, basic and aromatic amino acids, respectively. P_1 to P_{13} refer to the amino acid position within the sequence.

The screening of the optimized CPLPPXP library on Nck1 brought similar results to that obtained with the primary panning.

However, the affinity selection with Nck2 gave more information. Once more, no significant quantitative enrichment has been observed. However, randomly selected clones have been sequenced after the second round of affinity selection, this time some peptides other than the Non Proline ones were collected, which present the structure expected for SH3 domains, with the proline rich cassette. These clones show some homologies amongst themselves, as presented in Figure III-25.

A short overview on the proline rich clones bring evidence to the fact that the position directly after the proline rich motif, is highly advantageous for Nck2 SH3 domain, as an aromatic residue (mainly tryptophan, or tyrosine). But besides this position, the peptides appeared to be very different. However, a closer look allows the classification of the clones in three groups. The clones of the first group were gathered together, since they all show a defined proline on position P_{12}. The alignment of these clones highlights other homologies, like a common cysteine, or threonine on position P_3, or hydrophobic residues on positions P_4 and P_{13}. Furthermore, the position P_8 seems to be clearly defined as a phenylalanine.

The second group was formed with the clones presenting an arginine on position P_5. And here again, their alignment puts the accent on other similarities. Serines (or cysteines at P_{13}) might be of interest, on both extremities, and a leucine on position P_{11} also seems important.

The third set regroups the clones which did not belong to the first two classifications, and here as well, one can be surprised to realize that these clones show a very high content of hydrophobic residues.

Another point about the Nck clones, concerns the presence of cysteines. It is very uncommon to find clones against a particular target, with an odd number of cysteines. However, out of the 17 clones presented, 9 of them contain one cysteine residue (none of them possess a pair of cysteines). Since more than 50% of the clones (or 66% of the clones from the first two groups) show this characteristic, one can conclude that it is significant, even if the reason is unclear. Moreover, the calculated frequency of cysteines, within the CPLPPXP library is only 3,6% (Table III-10). Within the first group, 4 out of the 5 peptides have a cysteine, and this cysteine is always located before the proline rich motif (positions P_2 to P_4). As far as the second group is concerned, 4 out of the 7 clones also exhibit an unique cysteine, but this time, the cysteine is C terminal from the proline motif, or within the prolines (P_8 and

P_{13}). It is not known if this cysteine can build a disulfide bridge with a cysteine present in the Nck2 SH3 domain.

When clones from the third round were sequenced, the frequency of the NP1 and NP2 largely dominates the other clones.

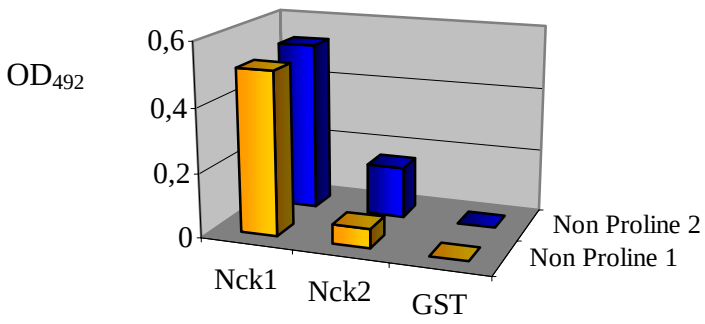
In the case of the Nck2 SH3 protein, the cosmix-plexing® recombination appears to be the only technology allowing isolation of proline rich clones. Until now, no laboratories report successful characterization of Nck ligands. In our work, the primary panning only permits the enrichment of non proline clones. The cosmix-plexing® recombination can therefore be considered as a second chance, for target proteins which would appear complex to investigate on affinity selection.

2-3 Phage ELISA

The clones characterized after the primary panning (NP1 and NP2), and the one obtained from cosmix-plexing® panning (Figure III-25), were propagated as single phage, and individually checked on phage ELISA for affinity against the Nck1 (NP1 and NP2) and Nck2 SH3 domains.

The wells of an ELISA microtiter plate were coated with the GST-Nck1 or-Nck2 proteins. A well, coated with GST alone, was used as control to ensure that the single phage interact with the Nck1 and Nck2 SH3 domains, and not with the GST. None of the clones isolated from the cosmix-plexing® panning show positive interaction for the Nck2 domain (not shown), which is not surprising, since none of the clones get enriched. The results are summarized in Figure III-26.

On phage ELISA, only the NP1 and NP2 clones react positively with the Nck1 and Nck2 domains. The non proline clones interact strongly with Nck1, but weakly with Nck2 SH3 domain. This is likely to be the reason why it has been possible to isolate some new clones for the Nck2 domain, since they had a better chance to compete with the non proline clones. These results, together with the fact that only the NP clones get enriched, demonstrate that the NP clones have a selective advantage on the proline rich clones, to bind to the Nck domain. The GST does not react with any of the clones, demonstrating their specificity for SH3 domains.



	Phage ELISA of Non Proline clones against Nck1 and Nck2 SH3 domains								OD_{492} Nck1	Nck2
NP1	Ser	Thr	Leu	**Cys**	**Asp**	**Gly**	**Tyr**	**Cys**	0,51	0,53
NP2	Asn	Pro	Met	Gly	**Asp**	**Gly**	**Tyr**	Phe	0,06	0,16
Consensus sequence	Xxx	Xxx	Xxx	Xxx	**Asp**	**Gly**	**Tyr**	Xxx		

Figure III-26: Relative affinity, according to a phage ELISA test, of the clones isolated against Nck1 and Nck2 SH3 domains. GST-Nck1 and -Nck2 proteins were coated in the wells. An equivalent amount of phage were incubated in every well (10^{10} phage/well). After several washes, the remained bound phage were detected with anti-M13 antibody. The absorbance was detected at 492 nm. A well coated with GST alone was used as a control, to ensure that the single clones bind to the SH3 domains, and not to the GST. The OD values are an average of three repeated tests.

3- Nephrocystin SH3 domain

Nephrocystin cDNA was obtained from Friedhelm Hildebrandt (University of Freiburg; Germany). Nephrocystin SH3 domain (amino acids 159 to 208 of the *NPHP1* cDNA gene, GenBank accession number AF023674) was expressed as GST-fusion protein, by cloning the PCR amplified fragments in vector pGEX-5X3, as previously described (Smith and Johnson, 1988). A highly purified GST-nephrocystin fusion protein was obtained (Figure III-2).

3-1 Panning selection

The purified GST-nephrocystin protein was used as target for affinity selection with the CPLPPXP library, according to the protocol optimized for Src SH3 domain (section III-D-1-1). Briefly, a maxisorb microtiter plate was used as support for the immobilization of the protein. After incubation of the nephrocystin with the CPLPPXP library, the non- or weak-binders were eliminated with mild washing steps, becoming more stringent at each cycle of panning (procedure A, Table III-13). The bound phage were eluted with acidic buffer, selected, and amplified for the next round of affinity selection. The evolution of input and output values can be followed in Figure III-27.

After three rounds of panning procedure, the enrichment factor appeared not to be significant (only 2.7, as shown in Figure III-27). However, if we consider the enrichment between the second and third round of panning, a 20 fold increase is detected. Therefore, 15 randomly selected clones were sequenced.

A

	Input	Output	Output/Input	Enrichment
1R	1,50E+10	1,30E+03	8,60E-08	-
2R	7,00E+10	9,00E+02	1,20E-08	-
3R	3,90E+10	9,30E+03	2,30E-07	2,7

B

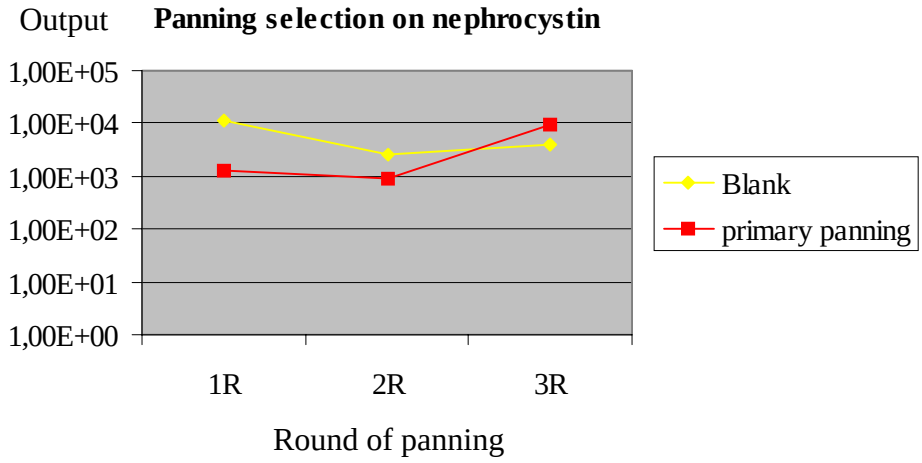


Figure III-27: Evolution of the input and output titers during the screening of CPLPPXP on nephrocystin SH3 domain. (A-) Progression of the input (amount of phage incubated with the target protein), output (eluted phage after stringent washing steps), and output/input (O/I) values, in absolute number, during 3 rounds of panning procedure (noted 1R to 3R). The O/I ratio is a parameter indicating if an enrichment occurred. The enrichment factor is estimated by dividing the ratio O/I from the round of interest, with the ratio O/I from the first round. (B-) Diagram of the output values of the primary panning on nephrocystin SH3 domain, compared with the output data of the blank. The blank is an empty well where the CPLPPXP library was incubated, and then, treated like in the other well.

Very few clones were recovered after the first panning. However, as a compensation, the usual dramatic drop of recovery observed after the second round, did not occur, instead, the output stabilized, and finally increased in the third round, showing an inverse tendency in comparison to the evolution of the control. This inverted trend suggests that interactions between the CPLPPXP library and the nephrocystin occurred.

The sequenced clones confirmed this hypothesis, since out of 15 clones, none of them expressed unproductive clones (change in the open reading frame or presence of stop codon). Furthermore, from these 15 peptides, only three different sequences were observed, which led to the conception of a clear consensus sequence (Figure III-28).

This consensus sequence shows very high homologies with the one conceived for the Src SH3 domain, as shown:

Nephrocystin **Arg** Xxx **Leu** **Pro Pro** Xxx **Pro** **Val** **Pro**

Src **Arg** **Pro** **Leu** **Pro** **P/L** **Ψ** **Pro** **Val** **Pro** **Pro**

The only position which really seems to be different, is the second one (or P_4 on complete sequences), which is defined as proline in the Src consensus sequence, but appears to be free for the nephrocystin.

	Sequences against Nephrocystin SH3 domain (3rd round of panning)														Frequency
	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}	P_{13}		
N1			**Arg**	Ser	**Leu**	**Pro**	Leu	Pro	**Pro**	**Val**	**Pro**	Thr	Ser	Ser	Ala	7x
N2	Ser	Val	**Arg**	Leu	**Leu**	**Pro**	**Pro**	Val	**Pro**	Trp	His	Phe	Gly			6x
N3	His	Ser	**Arg**	Gln	**Leu**	**Pro**	**Pro**	Lys	**Pro**	**Val**	**Pro**	Ser	Leu			2x
Consensus sequence	Xxx	Xxx	**Arg**	Xxx	**Leu**	**Pro**	**Pro**	Xxx	**Pro**	**Val**	**Pro**	Xxx	Xxx			

Figure III-28: Sequences of 15 randomly picked clones, after a third round of panning selection of the CPLPPXP library, against nephrocystin SH3 domain. The clones are classified according to their frequency. A residue is considered as representative for the consensus sequence, when it occurs in 50% or more of the sequences. Underlined residues are fixed within the CPLPPXP library. P_1 to P_{13} refer to the amino acid positions within the sequence.

3-2 Cosmix-plexing® recombination

After this successful panning, a cosmix-plexing® recombination was performed, in order to isolate new ligands against the nephrocystin SH3 domain. It was hoped that these would throw more light on the sequences which determine specific binding to nephrocystin SH3 domain.

Since only very few particles were recovered after the first round of panning, the clones of the first round population were recombined among themselves. The strategy used for this approach is described in Figure III-22. The nephrocystin secondary library was screened against the nephrocystin SH3 domain, and the input and output values, along with the enrichment factors are summarized in Figure III-29.

A

	Input	Output	Output/Input	Enrichment
1R	1,90E+09	3,00E+02	1,50E-07	-
2R	1,10E+10	5,40E+03	4,90E-07	3,2
3R	7,20E+10	6,20E+06	8,60E-05	574

B

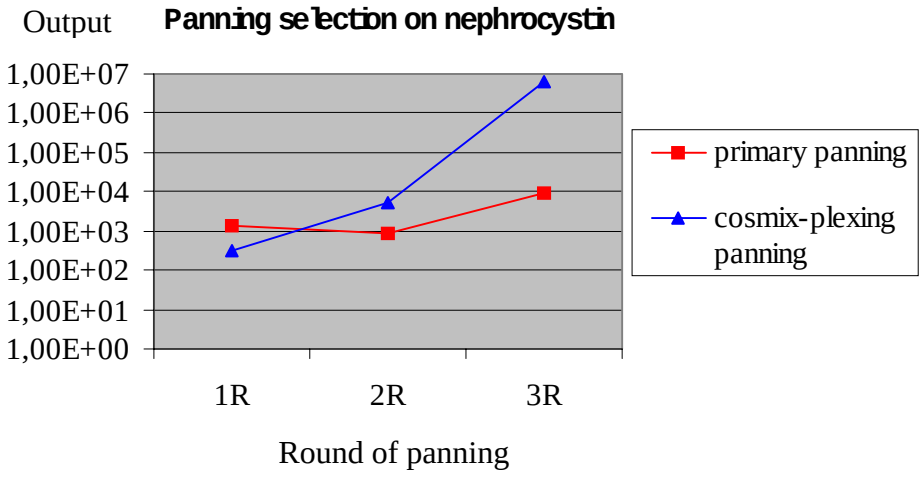


Figure III-29: Evolution of the input and output titers during the screening of the cosmix-plexed secondary nephrocystin CPLPPXP library. (A-) Progression of the input (amount of phage incubated with the target protein), output (eluted phage after stringent washing steps), and output/input (O/I) values, in absolute number, during 3 rounds of panning procedure (noted 1R to 3R), along with the enrichment factors. (B-) Diagram of the output values of the cosmix-plexing® panning on nephrocystin SH3 domain, compared with the output data observed for the primary panning of CPLPPXP library against nephrocystin (as shown in Figure III-27).

After the third panning cycle, a very clear enrichment was observed, since the enrichment factor was more than 200 fold higher than the one obtained during the primary affinity selection (574, compared to 2.7 during the primary panning). In the diagram (Figure III-29/B), the evolution of the output values of both primary and secondary pannings against the nephrocystin SH3 domain are compared. This illustrates a clear advantage of cosmix-plexing® recombination, in comparison to the initial library.

Randomly selected clones after the second, third and fourth rounds were sequenced (Figure III-30). The peptides were kept separate within the figure, so that the evolution of the enrichment is easier to follow.

The clones isolated during the primary panning did not reappear during the secondary panning. Many new clones were characterized, suggesting that they have higher binding affinity for nephrocystin than the N1, N2, and N3 clones.

It also has to be pointed out that most of the newly isolated clones do not contain the proline rich cassette. When the CPLPPXP library was constructed, only 71% of the vectors have been shown to possess a cassette (Figure III-11), while 20% of them formed back the parental vectors, containing the 8 random amino acid insert (the last 9% of the library generated unproductive clones, which formed the so-called L clones, Figure III-18).

It is known that the SH3 ligands are expected to be 6 or 7 amino acids long. Therefore, it is surprising, but not impossible that the 8 random amino acid peptides, have an advantage over the 13 residue proline rich peptides, to interact with nephrocystin. This would be the case especially if the Pro Pro Xxx Pro core motif was not optimal for nephrocystin. Furthermore, these 8 residue clones show a similar consensus sequence to the one obtained from the primary panning, indicating that it is not a coincidence if these clones were enriched. A proline core motif (Pro **Val** Xxx Pro) is present, as well as the preceding arginine and leucine.

Consensus sequence after primary panning

Arg Xxx **Leu** **Pro Pro** Xxx **Pro** **Val** **Pro**

Consensus sequence after secondary panning (cosmix-plexing®)

Arg Xxx **Leu** **Pro** **Val** Xxx **Pro**

	Sequences against Nephrocystin SH3 domain, after cosmix plexing	Frequency (%)
2nd round		
N4	Thr **Arg** Arg **Leu Pro Val** Leu **Pro**	58
N5	**Arg Met Leu Pro Val** Tyr **Pro Pro**	12
N6	***Arg*** Thr **Pro Pro** Thr Glu **Pro** Gln Thr	6
N7	**Arg Met Leu Pro Val** Lys Ile Tyr	6
N8	***Arg*** **Pro Leu Pro Pro** Thr **Pro Pro** Ile Pro Gly Ser Gly Ser	6
N9	Cys Ser Gln His Asp Pro Pro Phe Pro Trp Arg Val Cys	12
Consensus sequence 2nd round		
	Arg M/Ψ Leu Pro Val Xxx **Pro Pro**	
3rd round		
N4	Thr **Arg** Arg **Leu Pro Val Leu Pro**	37
N5	**Arg Met Leu Pro Val** Tyr **Pro** Pro	29
N10 Gly	Asn **Arg Pro Leu Pro Pro Ile Pro** Ser His Pro Phe	25
N11 Arg	Ala **Arg Val Val Pro Pro Val Pro** Arg Leu His Thr	3
N9	Cys Ser Gln His Asp Pro Pro Phe Pro Trp Arg Val Cys	6
Consensus sequence 3rd round		
	Xxx **Arg Ψ Leu Pro V/P Ψ Pro** Xxx	
4th round		
N4	Thr **Arg** Arg **Leu Pro Val** Leu **Pro**	30
N5	**Arg Met Leu Pro Val** Tyr **Pro** Pro	70
Consensus sequence 4th round		
	Xxx **Arg** Xxx **Leu Pro Val** Xxx **Pro** Xxx	

Figure III-30: Sequences of randomly picked clones, cosmix-plexing® affinity selection against nephrocystin SH3 domain. 17, 35 and 10 clones were sequenced, after the 2nd, 3rd, and 4th round of panning, respectively. The frequency of clone occurrence, is given in percentage, in order to easier estimate the evolution of the clone appearance after the different cycles of panning. Underlined residues are fixed within the CPLPPXP library. The first amino acid of both N6 and N8 sequences (Arg) are written in italic. These Arg do not belong to the hypervariable domain of the library, but to the conserved sequence which flanks the variable region. However, these residues seem to be of particular interest for the binding abilities of the clones for the Nephrocystin SH3 domain (by reference to the consensus sequenced), and therefore, it was decided to outline them. The N9 sequence was not taken into account when the consensus sequence was conceived. Indeed, it would not be right to compare it to the other clones, since the presence of cysteines on both extremities suggest the formation of a disulfide bridge, which might confer a very particular conformation to the clone. Ψ represents the hydrophobic residues.

The interesting point, is that the proline core motif is significantly different from the biased core sequence used in the construction of the library. It does not contain 3, but 2 prolines (PVXP, instead of PPXP, as the CPLPPXP library was designed). Therefore, a hypothesis would be that the nephrocystin required a proline motif designed as Pro Val Xxx Pro, and not the Pro Pro Xxx Pro necessary for the Src SH3 domain ligands. To acquire ligands containing this required proline motif, the nephrocystin interacts with the 20% of the CPLPPXP library which show no fixed proline, and selects the variants exhibiting two prolines separated by two other amino acids.

Very surprisingly, a peptide (N9) bracketed with two cysteines has been isolated, which occurred several times (in 12% of the sequenced clones, after the second round). This N9 peptide is expected to form a disulfide bridge, and therefore, to exhibit a conformation other than that found in unconstrained clones. As expected, N9 does not show any similarity to the other characterized clones. Since the presence of the cysteines can direct the formation of a loop, the structure of the ligands should be completely different from the other linear clones. Furthermore, it might interact with nephrocystin in another way, meaning that the contact between the ligand and the SH3 domain would be achieved by other residues (therefore, this clone was not taken into account in creating the consensus sequence). However, it was previously shown, that constrained libraries were not successful in panning to isolate SH3 ligands, presumably since the constrained peptides were not able to form the required polyproline type II helix. It was therefore unexpected to obtain such a clone. Further analysis will give more information about the interaction between N9 and nephrocystin.

After the second round, clone N4 was dominant (58%). However, after the third round, a equilibrium between N4 and N5 dominance was nearly achieved (37% against 29%). The fourth round revealed that the N5 clone is finally more advantageous for the interaction with nephrocystin, since it occurred then in 70% of the cases. However, after a fourth round, N4 and N5 are the only enriched clones, suggesting that they both show high affinity for nephrocystin.

3-3 Phage ELISA

The consensus sequence obtained after the cosmix-plexing® panning procedure (**Arg** Xxx **Leu** **Pro** **Val** Xxx **Pro**), appears very similar to the one conceived after the primary affinity selection (**Arg** Xxx **Leu** **Pro Pro** Xxx **Pro** **Val** **Pro**), except for the proline core motif which seems highly advantageous when containing only two prolines. However, these two

consensus sequences have been established from different clones. It is therefore essential to check the individual relative binding affinity of each clone, before making further conclusions.

Since the consensus sequence conceived for nephrocystin ligands is very similar to that established for Src ligands, it was decided to investigate cross reaction, on phage ELISA (Figure III-31). Furthermore, two identical clone types have been isolated in the selections against both nephrocystin and Src SH3 domains (N1 and N10 are respectively clones S3 and S1).

From the 11 clones isolated against nephrocystin, only the N9, which was bracketed by two cysteines, shows no interaction. It is however not clear if the N9 clone is really unable to bind nephrocystin, if it is too weak to show a binding on phage ELISA, or if it is dependent on redox potential.

A new consensus sequence is conceived from the clones showing high relative binding to nephrocystin, which is more extended, and containing less ambiguities than the one obtained from simple sequence comparisons.

Nephrocystin consensus sequence after primary panning

Arg Xxx **Leu** **Pro Pro** Xxx **Pro** **Val** **Pro**

Nephrocystin consensus sequence after cosmix-plexing® panning

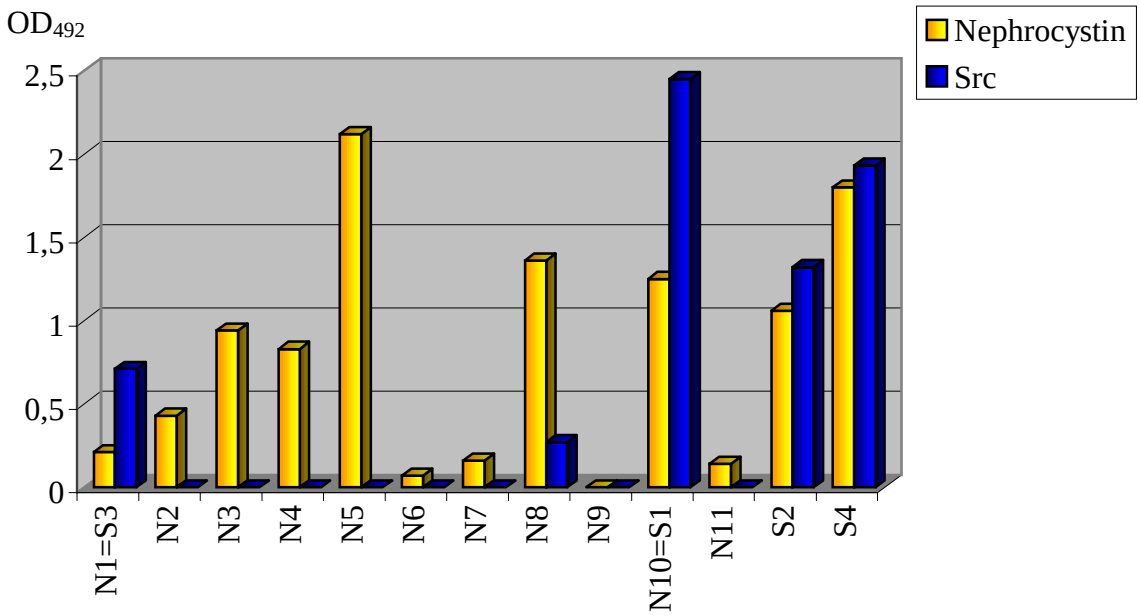
Arg Xxx **Leu** **Pro** **Val** Xxx **Pro**

Nephrocystin consensus sequence from positive clones after phage ELISA

β Xxx **Arg** **Pro** **Leu** **Pro** **P/Ψ** **(Ψ)** **Pro** **P/V** Xxx **Pro**

Figure III-31: Relative binding of nephrocystin, and Src positive clones to nephrocystin and Src SH3 domains. Isolated clones against Nephrocystin and Src SH3 domains were propagated as single phage, and tested on phage ELISA, against immobilized Nephrocystin and Src SH3 domains (1µg). 10^{10} phage of the single clones were incubated with the SH3 domains (except for the clone N5 were only 5.10^{8} phage were applied), after blocking the wells. The phage were detected after intensive washing, by antibodies against M13KO7, labeled with Peroxidase. A well coated with GST alone was used as a control, to ensure that the single clones bind to the Src, and not to the GST. The phage showed no interaction with the GST (not shown). The sequence of the clones which react positively with the nephrocystin domain are shown, and a new consensus sequence is conceived, from the clones presenting a high absorbance. The OD values are an average of three repeated tests. Underlined residues are fixed within the CPLPPXP library. Residues in italic represent positions which do not belong to the hypervariable peptide, but to the flanked regions. These residues were considered as important, due to their similarity to the consensus sequence. Ψ and β represent the hydrophobic and basic residues, respectively.

Phage ELISA on nephrocystin and Src SH3 domains



	Positive clones against Nephrocystin SH3 domain,															OD492
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13			
Strong binding clones																
N5			**Arg**	**Met**	Leu	**Pro**	**Val**	Tyr	**Pro**	**Pro**	*Asp*	***Pro***				2,1
S4	**Lys**	Gly	**Arg**	**Pro**	Leu	**Pro**	**Pro**	**Val**	**Pro**	Gly	Thr	**Pro**	Ser			1,8
N8			***Arg***	**Pro**	Leu	**Pro**	**Pro**	Thr	**Pro**	**Pro**	Ile	**Pro**	Gly	Ser	Gly	1,3
N10/S1	Gly	Asn	**Arg**	**Pro**	Leu	**Pro**	**Pro**	**Ile**	**Pro**	Ser	His	**Pro**	Phe			1,2
S2			**Arg**	**Pro**	Leu	**Pro**	**Leu**	**Pro**	**Pro**	**Val**	Pro	Trp	Val	Arg	Trp	1
N3	**His**	Ser	**Arg**	Gln	Leu	**Pro**	**Pro**	Lys	**Pro**	**Val**	Pro	Ser	Leu			0,9
N4	***Arg***	Thr	**Arg**	Arg	Leu	**Pro**	**Val**	**Leu**	**Pro**	*Asp*	*Pro*					0,8
Consensus sequence																
	β	Xxx	**Arg**	**Pro**	Leu	**Pro**	**P/Ψ**	**(Ψ)**	**Pro**	**P/V**	Xxx	**Pro**	Xxx			
Weak binding clones																
N2	Ser	Val	**Arg**	**Leu**	Leu	**Pro**	**Pro**	**Val**	**Pro**	Trp	His	Phe	Gly			0,4
N1/S3			**Arg**	Ser	Leu	**Pro**	**Leu**	**Pro**	**Pro**	**Val**	Pro	Thr	Ser	Ser	Ala	0,2
N6			***Arg***	Thr	**Pro**	**Pro**	Thr	Glu	**Pro**	Gln	Thr					0,1
N7			**Arg**	**Met**	Leu	**Pro**	**Val**	Lys	Ile	Tyr						0,1
N11	**Arg**	Ala	**Arg**	**Val**	**Val**	**Pro**	**Pro**	**Val**	**Pro**	Arg	Leu	His	Thr			0,1

The proline core motif does not require three prolines. Two prolines (P_6 and P_9) separated by two hydrophobic residues, are enough to give excellent binding ability to the nephrocystin ligands. The same goes for the proline between the arginine and the leucine (P_4), 4 out of the 7 clones which positively react with nephrocystin exhibit a proline on this position. However, the best clone (N5), do not possess it, suggesting that this proline can present an advantage, but is not crucial.

The comparison of the two consensus sequences also show that the one established after phage ELISA gives more information on the flanking regions, N and C terminal from the proline core motif.

The analysis of the clones which bind very weakly to the nephrocystin (N6, N7, and N11), also provides a lot of information in determining which positions appear to be crucial for the interaction. All three clones differ only by one or two positions from the basic consensus sequence. N6 does not have Leu_5 and Ψ_7. N7 is missing Pro_9, and N11 lacks the Leu_5 position. These data allow a more stringent definition of the crucial nephrocystin ligand consensus sequence:

Arg Xxx **Leu** **Pro** **Ψ** Xxx **Pro**

The four clones isolated against Src (two were also selected against nephrocystin) are able to interact with both nephrocystin and Src SH3 domains. Figure III-32 illustrates the sequences of the interacting clones. However, of the other nephrocystin clones, only N8 shows a cross reaction with Src. Every clone able to bind to Src, in our phage ELISA test, is also able to bind to nephrocystin, but the contrary is not true. The ligand preference of Src appears to be more demanding than that of Nephrocystin.

The analysis of the nephrocystin clones which do not interact with Src, clearly show an absence of the Arg Pro Leu motif preceding the proline rich core. This again demonstrates the crucial role of this Arg Pro Leu sequence in Src SH3 ligands, as previously pointed out by the work of other groups and confirmed here (Figure III-23).

The new Src consensus sequence, conceived from the totality of the clones able to interact with Src (no matter if isolated against Src or nephrocystin) does not significantly alter the previously defined consensus sequence (since only one more clone is participating in its elaboration), however, one position more is defined.

Src consensus sequence after Src phage ELISA:

Arg Pro Leu Pro P/L Ψ Pro Val Pro Pro S/Ψ

Src consensus sequence by considering all interacting clones

Arg Pro Leu Pro P/L Ψ Pro Ψ Ψ Pro S/Ψ Ser

Positive clones against Src SH3 domain, ordered according to their relative affinity for the target															**OD_{492}**
N10/S1	Gly	Asn	**Arg**	**Pro**	**Leu**	**Pro**	**Pro**	**Ile**	**Pro**	Ser	His	**Pro**	**Phe**		2,4
S4	Lys	Gly	**Arg**	**Pro**	**Leu**	**Pro**	**Pro**	**Val**	**Pro**	Gly	Thr	**Pro**	**Ser**		1,9
S2			**Arg**	**Pro**	**Leu**	**Pro**	**Leu**	**Pro**	**Pro**	**Val**	**Pro**	**Trp**	**Val**	Arg	Trp	1,3
N1/S3			**Arg**	Ser	**Leu**	**Pro**	**Leu**	**Pro**	**Pro**	**Val**	**Pro**	Thr	**Ser**	**Ser**	Ala	0,7
N8			***Arg***	**Pro**	**Leu**	**Pro**	**Pro**	Thr	**Pro**	**Pro**	**Ile**	**Pro**	Gly	**Ser**	Gly	0,3
Consensus sequence															
	Xxx	Xxx	**Arg**	**Pro**	**Leu**	**Pro**	**P/L**	**Ψ**	**Pro**	**Ψ**	**Ψ**	**Pro**	**S/Ψ**	**Ser**	Xxx

Figure III-32: Sequences of the clones isolated against nephrocystin or Src SH3 domains, and which interact with Src SH3 domains on phage ELISA. A new consensus sequence is conceived, from the clones presenting relative binding. The OD values are given, as an average of three repeated tests. Underlined residues are fixed within the CPLPPXP library. Residues in italic represent positions which do not belong to the hypervariable peptide, but to the flanked regions. These residues were considered as important, due to their similarity to the consensus sequence. Ψ represents the hydrophobic residues.

The C terminal region of the consensus sequence, shows a few differences which are, however, compatible to the previous consensus sequence (the Val_{10} and the Pro_{11} are both hydrophobic residues). A serine at position P_{14} appears now to be advantageous for the binding between Src SH3 domain, and its ligands.

E - Panning selection on EVH1 domains

The Ena/VASP homology (EVH1) domain is a protein interaction module found in several *Drosophila* and mammalian proteins (the *Drosophila* protein Ena, their murine homologue proteins Mena and Evl, the human proteins VASP, WASP and Vesl/Homer) that are involved in transducing migratory and morphological signals into cytoskeletal reorganization (Prehoda et al., 1999). EVH1 specifically recognizes proline-rich sequences in

its binding partners. They direct the localization and formation of multicomponent assemblies involved in actin-biased motile processes by recruiting proteins like profilin and the Arp2/3 complex, which are thought to directly promote actin filament elongation (Mitchison and Cramer, 1996; Laurent et al., 1999). In addition, Vesl is involved in neural development (Tu et al., 1999).

Mena, VASP, Evl, and Vesl EVH1 domains, were kindly provided as purified GST-fusion proteins, by Prof. J. Wehland (GBF, Braunschweig, Germany).

1- Vesl/Homer EVH1 domain

Vesl, also called Homer, belongs to a family of postsynaptic proteins that are thought to function in synaptic plasticity (Brakeman et al., 1997; Kato et al., 1997). Vesl is constitutively expressed in brain and enriched at excitatory synapses, that selectively binds the C terminus of group 1 metabotropic receptors (mGluR1a and mGluR5) (Brakeman et al., 1997). Vesl interacts with mGluR1a/5 through its EVH1 domain. The Vesl EVH1 domain binds an internal, proline rich sequence that is approximately 50 residues from the C terminus of both mGluR1a and mGluR5 (Tu et al., 1998).

Although Vesl contains a EVH1 domain, it does not seem to be involved in actin cytoskeleton, but to function in synaptic plasticity (Brakeman et al., 1997; Kato et al., 1997). This element demonstrates that Vesl stands apart from the other EVH1 domain containing proteins. This is also illustrated by the ligand preferences of EVH1 domains, as shown in Figure III-33. There, the natural EVH1 ligands are divided into two groups. EVH1 domains involved in actin filament assembly show predilection for the first set (ActA, vinculin, zyxin, ankyrinG, SAX-3 and Robo), whereas Vesl binds preferentially to the second group, composed by mGluR, and IP3.

Vesl is the only EVH1 domain containing protein, to interact with the proline rich motif Pro Pro Xxx Pro Phe Arg Asp, found in mGluR and IP3. In addition, whereas all other EVH1 domains show strong affinity to the ActA protein, Vesl interacts very weakly.

The main difference between both group of ligands, is the position of the phenylalanine, before (first group), or after (second group) the proline core motif. Phe directly before the proline motif, is essential for the binding with EVH1 domains, since interaction of ActA with EVH1 domain is disrupted, when Phe is mutated to Ala (Prehoda et al., 1999). EVH1 domains contain a pocket that specifically recognizes this Phe, or large hydrophobic residues, such as leucine, in the human Robo sequence.

Robo does not contain Phe, either before or after the core motif, but in both cases, other large hydrophobic amino acids. This might explain a certain loss of affinity for the binding with EVH1, compared to the other ligands. Furthermore, Robo shows an interaction with both EVH1 domain families, even though quite weak, which might be possible through the absence of the phenylalanine which seems to direct the specificity for one of the two classes of EVH1 domains.

Natural ligands of EVH1 domains

ActA		Ser	Phe	Glu	Phe	Pro	Pro	Pro	Pro	Thr	Asp	Glu	Glu	Leu
Vinculin		Glu	Pro	Asp	Phe	Pro	Pro	Pro	Pro	Pro	Asp	Leu	Glu	Gln
Zyxin	Glu	Glu	Ile	Phe	Phe	Ser	Pro	Pro	Pro	Pro	Pro	Glu	Glu	Glu
AnkyrinG	Ala	Tyr	Ile	Glu	Phe	Pro	Pro	Pro	Pro	Pro	Leu	Asp	Ala	Asp
SAX-3	Thr	Leu	Met	Asp	Phe	Ile	Pro	Pro	Pro	Pro	Ser	Asn	Pro	Pro
hRobo	Ser	Thr	Asp	Asp	Leu	Pro	Pro	Pro	Val	Pro	Pro	Pro	Ala	Ile
Consensus sequence														
	Xxx	Xxx	Ψ	α	Phe	Pro	Pro	Pro	Pro	Pro	Ψ	α	Glu/Ψ	Ψ

mGluR	Glu	Glu	Leu	Val	Ala	Leu	Thr	Pro	Pro	Ser	Pro	Phe	Arg	Asp
IP3R							Asn	Pro	Pro	Lys	Lys	Phe	Arg	Asp
Dynamin III							Ala	Pro	Pro	Val	Pro	Phe	Arg	Pro
Shank3					Pro	Val	Pro	Pro	Pro	Glu	Glu	Phe	Ala	Asn
Consensus sequence														
							Xxx	Pro	Pro	Xxx	Pro	Phe	Arg	Asp

Figure III-33: Alignment of natural EVH1 binding sequences, including sequences from the *Listeria* ActA protein, host cytoskeletal proteins (vinculin, zyxin, ankyrin), and host axon guidance receptors (SAX-3, Robo). mGlu receptor and IP3 receptor, are ligands showing interaction specifically for Vesl, but not with the other EVH1 domains. They are therefore kept separated from the previous group of ligands. Deduced consensus sequences are shown. α and Ψ represent acidic and hydrophobic amino residues, respectively.

Another important peculiarity of EVH1 ligands, is the presence of an acidic stretch, flanking both sides of the proline motif, which appears to increase the affinity of the ligands for the EVH1 domains (Niebuhr et al., 1997; Prehoda et al., 1999).

The actual consensus sequence for EVH1 domains involved in actin filament assembly, or in synaptic plasticity, at the beginning of this work, were conceived from the natural EVH1 ligands(Figure III-33), and were as followed:

Actin filament assembly	Ψ	α	**Phe**	**Pro**	**Pro**	**Pro**	**Pro**	**Pro**	Ψ	α	**Glu**/Ψ	Ψ
Synaptic plasticity				**Pro**	**Pro**	Xxx	**Pro**	**Phe**	**Arg**	**Asp**		

It has to be noticed that acidic residues do not necessarily occur within the consensus sequences, since from one ligand to another, the many acidic positions seem to vary all along the ligand sequence.

1-1 Panning selection

The purified GST-Vesl fusion protein, was submitted to affinity selection, according to the optimized protocol developed for the Src SH3 domain. A maxisorb microtiter plate was used as support for the immobilization of the protein. After incubation of Vesl with the CPLPPXP library, the non- or weak-binders were eliminated by proceeding to mild washing steps, becoming more stringent at each cycle of panning (procedure A, Table III-13). The bound phage were eluted with acidic buffer, selected, and amplified for the next round of affinity selection. The evolution of the input and output values can be followed in Figure III-34.

After three rounds of panning procedure, the enrichment factor is very weak (only 5,5, as shown in Figure III-34). However, after the fourth round, a dramatic increase is observed, with an enrichment factor of 600

The usual drop of recovery is, as expected, observed after the second round. The output values rise slightly in the third round, and finally increased significantly in the fourth round, in comparison to the control, which shows no increase.

Randomly selected clones from the third round were sequenced. A strong enrichment of particular clones occurred: out of 25 sequenced clones, only one of them was an unproductive clone (presence of a stop codon). Furthermore, among the 24 correctly formed peptides, several of them occurred several times, and their alignment leads to the conception of a clear consensus sequence (Figure III-35).

A

	Input	Output	Output/Input	Enrichment
1R	1,50E+10	1,80E+04	1,20E-06	-
2R	6,00E+10	1,90E+03	3,10E-08	-
3R	6,00E+09	4,00E+04	6,60E-06	5,5
4R	3,8.9	2,40E+06	7,20E-04	600

B

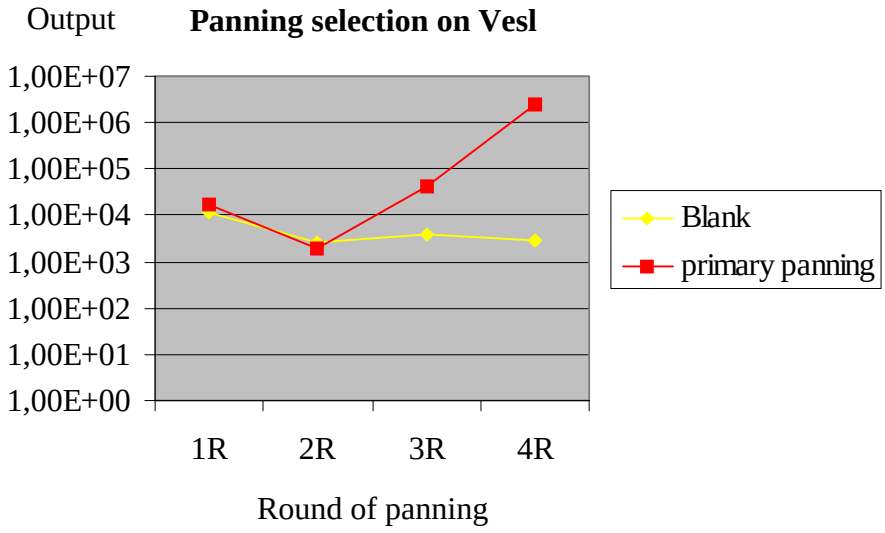


Figure III-34: Evolution of the input and output titers during the screening of CPLPPXP on Vesl EVH1 domain. (A-) Progression of the input (amount of phage incubated with the target protein), output (eluted phage after stringent washing steps), and output/input (O/I) values, in absolute number, during 4 rounds of panning procedure (noted 1R to 4R). The O/I ratio is a parameter indicating if an enrichment occurred. The enrichment factor is estimated by dividing the ratio O/I from the round of interest, with the ratio O/I from the first round. (B-) Diagram of the output values of the primary panning on Vesl EVH1 domain, compared with the output data of the blank. The blank is an empty well blocked with the blocking solution, where the CPLPPXP library was incubated, and then, treated like in the other well.

The analysis of these clones shows the dominance of a tyrosine or tryptophan after the proline rich motif, at position P_{10}, and indicates a preference for Arg on position P_2 (only 50% of the clones present this Arg, however, these are all the clones which are clearly enriched. At position P_8, between the prolines, there is a clear preference for hydrophobic residues (or eventually for aromatics), as well as at the positions 3 and 4.

Clones sequenced after a fourth round of panning, demonstrate a further enrichment for clone V1 (90%) and V3 (10%). This reinforces the aptness of the conceived consensus sequence.

	Sequences against Vesl EVH1 domain (3rd round of panning)												Frequency
	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}	P_{13}	
V1	Asp	**Arg**	His	**Tyr**	Arg	**Pro**	**Pro**	**Phe**	**Pro**	**Trp**	Ala	**Asp**	Gly	13x
V2	Asn	**Arg**	**Leu**	**Tyr**	Pro	**Pro**	**Pro**	**Trp**	**Pro**	**Tyr**	Ser	**Asp**	Pro	4x
V3	Ser	**Arg**	Ser	**Val**	Tyr	**Pro**	**Pro**	**Pro**	**Pro**	**Tyr**	Pro	**Phe**	Ala	2x
V4	Asp	**Arg**	**Ile**	**Tyr**	Leu	**Pro**	**Pro**	**Val**	**Pro**	**Trp**	Ala	Asn	Ser	1x
V5	Arg	Thr	Gly	**Val**	Trp	**Pro**	**Pro**	**Pro**	**Pro**	His	Asp	**Phe**	Arg	1x
V6	Tyr	Leu	**Ile**	**Leu**	Ser	**Pro**	**Pro**	**Ala**	**Pro**	**Trp**	Arg	**Asp**	Arg	1x
V7	Cys	Ser	Arg	Ser	Arg	**Pro**	**Pro**	**Val**	**Pro**	Leu	Gly	Pro	Phe	1x
V8	His	Met	**Leu**	**Phe**	Asp	**Pro**	**Pro**	**Phe**	**Pro**	**Tyr**	Ser	Asn	Glu	1x
Consensus sequence	Xxx	**Arg**	**Ψ**	**(φ/Ψ)**	Xxx	**Pro**	**Pro**	**(φ/Ψ)**	**Pro**	**W/Y**	Xxx	**D/F**	Xxx	

Figure III-35: **Sequences of 25 randomly picked clones, after a third round of panning selection of the CPLPPXP library, against Vesl EVH1 domain** (one clone contained a stop codon). Underlined residues are fixed within the CPLPPXP library. φ and Ψ represent the aromatic and hydrophobic residues, respectively.

1-2 Cosmix-plexing® recombination

The population recovered after the first round of "normal" panning, was submitted to a cosmix-plexing®, as described in Figure III-22, and screened once more against Vesl. Randomly selected clones were sequenced after the first round of panning, since the number of eluted phage was already so high. The output and input data are shown in Figure III-36, and the sequenced clones are presented in Figure III-37.

After the third panning cycle, a dramatic enrichment was observed, since the enrichment factor is almost 200 fold higher than the one obtained during the primary affinity selection (1000, compared to 5,5 during the primary panning). In the diagram (Figure III-36/B), the evolution of the output values of a control, and both primary and secondary pannings against the Vesl EVH1 domain are compared. A very large number of clones are recovered after the first round of cosmix-panning, especially when compared with the data coming from the standard panning, meaning that the cosmix-plexing® recombination allowed the formation of many variants showing affinity for Vesl. During the following affinity selection cycles, the number of eluted clones regularly increase.

A

	Input	Output	Output/Input	Enrichment
1R	2,30E+12	1,00E+06	4,30E-07	-
2R	2,70E+11	5,00E+06	1,80E-05	42
3R	8,60E+10	3,70E+07	4,30E-04	1000

B

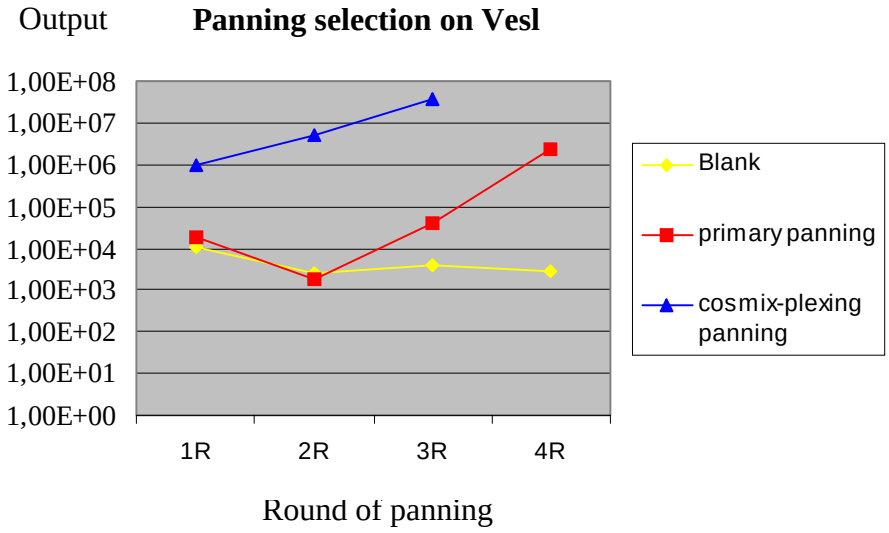


Figure III-36: Evolution of the input and output titers during the screening of the cosmix-plexed secondary Vesl CPLPPXP library. (A-) Progression of the input (amount of phage incubated with the target protein), output (eluted phage after stringent washing steps), and output/input (O/I) values, in absolute number, during 3 rounds of panning procedure (noted 1R to 3R), along with the enrichment factors. (B-) Diagram of the output values of the cosmix-plexing® panning on Vesl EVH1 domain, compared with the output data observed for the blank, and for the primary panning of CPLPPXP library on Vesl (as shown in Figure III-34).

The first element to be observed after cosmix-plexing®, with respect to the Vesl protein, is that it confirms the results found after the normal panning: the dominant clone is the same in both cases. This is also emphasized by the clones sequenced after the second round, since 100% of them were the V1 clone.

However, the cosmix-plexing® approach allows the obtention of a longer consensus sequence (only one of the 13 positions is still completely undetermined). In addition, less ambiguities within the consensus sequence are found. The tyrosine or tryptophan after the proline core motif, in position P_{10}, is confirmed, along with the P_2 Arg. The residue in P_8, between the prolines, shows a preference for a phenylalanine, or a proline, rather than just a hydrophobic residue (as concluded after the standard protocol). The fourth position is clearly determined as tyrosine, or eventually phenylalanine residue (but in any case, an aromatic

residue), which was not obvious after the normal panning.

	Sequences against Vesl EVH1 domain, after cosmix plexing (first round)												Frequency
	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}	P_{13}	
V1	Asp	**Arg**	His	**Tyr**	Arg	**Pro**	**Pro**	**Phe**	**Pro**	**Trp**	Ala	Asp	Gly	6x
V3	**Ser**	**Arg**	Ser	Val	Tyr	**Pro**	**Pro**	**Pro**	**Pro**	**Tyr**	Pro	**Phe**	**Ala**	1x
V9	**Ser**	**His**	**Ile**	**Phe**	**Asp**	**Pro**	**Pro**	**Phe**	**Pro**	**Tyr**	Gly	**Pro**	**Met**	1x
V10	**Ser**	Leu	**Leu**	**Tyr**	**Glu**	**Pro**	**Pro**	**Pro**	**Pro**	**Trp**	Asn	Ser	**Pro**	1x
V11	**Ser**	**Arg**	**Val**	**Tyr**	Pro	**Pro**	**Pro**	**Phe**	**Pro**	**Trp**	Arg	**Ala**	**Val**	1x
Consensus sequence														
	Ser	**Arg**	**Ψ**	**Y/F**	**(α)**	**Pro**	**Pro**	**F/P**	**Pro**	**W/Y**	Xxx	**Ψ**	**Ψ**	

Figure III-37: Sequences of 10 randomly picked clones, after a first round of panning selection. The CPLPPXP population recovered after the first round of ”normal” panning against the Vesl EVH1 domain, was submitted to cosmix-plexing®, and screened in this panning. Underlined residues are fixed within the CPLPPXP library. Ψ represents hydrophobic residues.

Furthermore, the cosmix-plexing® approach gives more information about both extremities of the consensus sequence. Indeed, the first position seems to be important as a serine. The most frequent clone (V1) does not contain this serine, however, the four other isolated variants do. Nevertheless, it has to be remembered that natural ligands show a quite high proportion of acidic residues, throughout the sequence, and that the V1 is the only clone isolated (after cosmix-plexing® recombination) which contains two acidic residues. One of these residues is in P_1 position, and therefore, might compensate the absence of serine. The last two positions (P_{12}, and $_{13}$) also appear to be preferentially hydrophobic residues.

The comparison of consensus sequences after standard, and cosmix-panning testifies an improved clarification of the consensus sequence, on the basis of the results obtained after cosmix-plexing®.

Consensus sequence after primary panning

Arg Ψ (φ/Ψ) Xxx **Pro Pro (φ/Ψ) Pro W/Y** Xxx **D/F**

Consensus sequence after secondary panning (cosmix-plexing®)

Ser Arg Ψ Y/F (α) Pro Pro F/P Pro W/Y Xxx **Ψ Ψ**

1-3 Phage ELISA

The clones characterized after the panning (and shown in Figures III-35 and-37), were prepared as single phage, and individually checked on phage ELISA for affinity against the Vesl protein. The results are summarized in Figure III-38.

The wells of an ELISA microtiter plate were coated with the GST-Vesl protein. A well, coated with GST alone, was used as control to ensure that the single phage interact with the Src, and not with the GST (not shown).

On phage ELISA, 10 out of the 11 isolated clones react positively with the Vesl domain, although some of them react very weakly. The alignment of the clones which react strongly with the Vesl EVH1 domain, allows the conception of a new consensus sequence, corresponding to the actual ligand preferences of Vesl:

Arg Ψ Y/F (Asp) Pro Pro F/Ψ Pro W/Y Xxx **Asp Ψ**

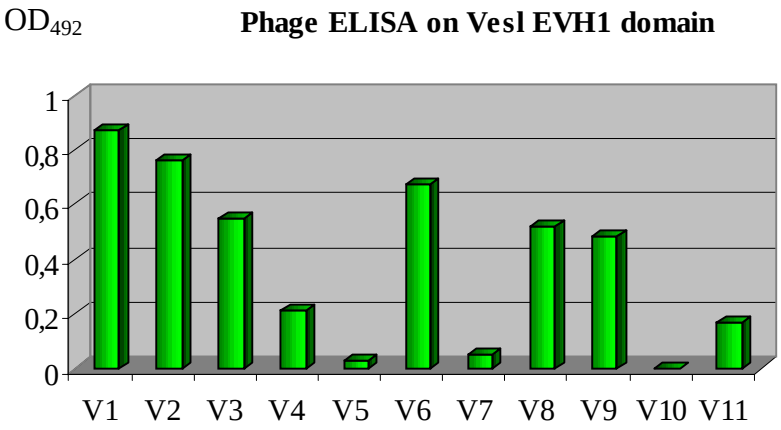
The analysis of the positive clones, shows how essential are the tryptophan or tyrosine at P_{10}, directly after the proline core motif.

In addition, it brings evidence for the importance of the arginine at position P_2, and tyrosine or phenylalanine at P_4. On both case, hydrophobic residues might be acceptable as well.

An aspartic acid at position P_{12} seems to strengthen the interaction, but not to be crucial, since the three best clones (V1, V2, and V6) contain it, but the others do not.

At P_8, phenylalanine, or hydrophobic residues belong to the ligand preferences of Vesl. In addition, hydrophobic residues at P_3 and P_{13} appears to be advantageous.

It has to be pointed out, that less acidic residues occur in the phage display isolated ligands, compared to the natural ligands, but in both cases, they are distributed along the sequence. However, the analysis of the positive clones tends to show that they could compensate the absence of one of the important positions (like P_2 or P_4). This might be due to stabilization of the ligand/Vesl complex, by an interaction with basic residues of the Vesl domain. The most strongly binding ligand (V1) contains two acidic residues.



	Phage ELISA on Vesl EVH1 domain (positive clones)												OD_{492}
	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}	P_{13}	
V1	Asp	**Arg**	His	**Tyr**	Arg	**Pro**	**Pro**	**Phe**	**Pro**	**Trp**	Ala	**Asp**	Gly	0,87
V2	Asn	**Arg**	**Leu**	**Tyr**	Pro	**Pro**	**Pro**	**Trp**	**Pro**	**Tyr**	Ser	**Asp**	**Pro**	0,76
V6	Tyr	Leu	**Ile**	**Leu**	Ser	**Pro**	**Pro**	**Ala**	**Pro**	**Trp**	Arg	**Asp**	Arg	0,67
V3	Ser	**Arg**	Ser	**Val**	Tyr	**Pro**	**Pro**	**Pro**	**Pro**	**Tyr**	Pro	Phe	**Ala**	0,55
V8	His	Met	**Leu**	**Phe**	Asp	**Pro**	**Pro**	**Phe**	**Pro**	**Tyr**	Ser	Asn	Glu	0,52
V9	Ser	**His**	**Ile**	**Phe**	Asp	**Pro**	**Pro**	**Phe**	**Pro**	**Tyr**	Gly	Pro	**Met**	0,48
Consensus sequence	Xxx	**Arg**	**Ψ**	**Y/F**	**(Asp)**	**Pro**	**Pro**	**F/Ψ**	**Pro**	**W/Y**	Xxx	**Asp**	**Ψ**	
V4	Asp	**Arg**	**Ile**	**Tyr**	Leu	**Pro**	**Pro**	**Val**	**Pro**	**Trp**	Ala	Asn	Ser	0,21
V11	Ser	**Arg**	**Val**	**Tyr**	Pro	**Pro**	**Pro**	**Phe**	**Pro**	**Trp**	Arg	Ala	**Val**	0,17
V7	Cys	Ser	Arg	Ser	Arg	**Pro**	**Pro**	**Val**	**Pro**	Leu	Gly	Pro	**Phe**	0,05
V5	Arg	Thr	Gly	**Val**	Trp	**Pro**	**Pro**	**Pro**	**Pro**	His	Asp	Phe	Arg	0,03

Figure III-38: Relative affinity, according to a phage ELISA test, of the clones isolated against Vesl EVH1 domain. GST-Vesl protein was coated in the wells. An equivalent amount of phage were incubated in every well (10^{10} phage/well). After several washes, the remained bound phage were detected with anti-M13 antibody. The absorbance was detected at 492 nm. A well coated with GST alone was used as a control, to ensure that the single clones bind to the Vesl, and not to its GST fusion partner. The phage showed no interaction with the GST (not shown). The sequence of the clones which react positively with the Vesl domain are shown, and classified according to their OD value. A new consensus sequence is conceived from the strongest clones. The OD values are the average value from triplicates. Underlined residues are fixed within the CPLPPXP library. Ψ represents hydrophobic residues.

2- Evl, Mena and VASP EVH1 domains

The EVH1 domain mediates the interaction of VASP, Mena and Evl (Ena-VASP-like), with cytoskeleton-interacting proteins, including the eukaryotic proteins vinculin and zyxin, and ActA of the motile intracellular bacterial pathogen *Listeria monocytogenes* (Niebuhr et al., 1997; Laurent et al., 1999).

VASP (vasodilator-stimulated phosphoprotein), is 380 amino acid human protein, and is a substrate of cyclic AMP-dependent and cyclic-GMP-dependent protein kinases (Halbrügge and Walter, 1989). It is found in a variety of tissues, and is associated with focal adhesions, microfilaments, and cell-cell contacts. VASP has been shown to be a multi-ligand protein that binds directly to filamentous actin (Reinhard et al., 1992), profilin (Reinhard et al., 1995a), the *Listeria monocytogenes* surface protein ActA (Chakraborty et al., 1995), zyxin (Reinhard et al., 1995b), and vinculin (Brindle et al., 1996; Reinhard et al., 1996). The binding of VASP to vinculin, zyxin, and ActA has been attributed to proline rich sequence from the ligands, which interacts with the VASP EVH1 domain (Niebuhr et al., 1997).

Mena and Evl are murine proteins that were identified by their similarity to *Drosophila* Enabled (Gertler et al., 1996). Brain extract analyses, from Mena knockout mice, have shown that the Evl protein was sufficient to support movement in the absence of Mena. Further studies brought evidence that Mena, Evl, and VASP play interchangeable roles in *Listeria* movement and that they most likely bind to the same molecular targets in this function (Laurent et al., 1999).

Many tissues contain two or three of the Mena, Evl or VASP proteins (Lanier et al, 1999). However, particular tissues contain them at different concentrations. Platelets are rich in actin-binding proteins involved in motility and contain micromolar amount of VASP, whereas Mena and Evl are present in low quantity (Aszodi et al., 1999). Brain extracts are rich in the Mena and Evl proteins, where VASP is expressed at low concentration.

2-1 Panning selection

The purified GST-Evl, -Mena, and -VASP proteins were submitted to affinity selection, according to the optimized protocol selected with Src SH3 domain. Briefly, a maxisorb microtiter plate was used as support for the immobilization of the protein. After incubation of the nephrocystin with the CPLPPXP library, the non- or weak-binders were eliminated with mild washing steps, becoming more stringent at each cycle of panning (procedure A, Table III-13). The bound phage were eluted with acidic buffer, selected, and amplified for the next round of affinity selection.

No quantitative enrichment was observed during the standard panning procedure. However, some clones were sequenced, and show interesting homologies (Figure III-39), even though each of the isolated clones occurred only once. It has to be emphasized that according to the proteins, and the rounds of panning, between 40 to 70% of the sequenced clones were leading to unproductive variants (change in reading frame, presence of stop codons, or L clones, sequences not shown). Furthermore, within the correctly formed variants, the non proline clones identified against Nck1 and Nck2 SH3 domains, were isolated several times.

Target proteins	Sequences											
	P_1	P_2	P_3	P_4	P_5	P_6 P_7	P_8	P_9	P_{10}	P_{11}	P_{12}	P_{13}
Evl	**Phe**	Trp	**Ile**	**Phe**	Ser	**Pro Pro**	**Val**	**Pro**	**Tyr**	Leu	Tyr	Ser
	Phe	Ile	**Val**	Arg	**Val**	**Pro Pro**	**Leu**	**Pro**	Leu	Thr	Thr	Asn
	Asp	Thr	**Ala**	**Val**	Asp	**Pro Pro**	**Phe**	**Pro**	**Trp**	Gly	Pro	Ile
	Trp	Ser	Gln	Arg	**Met**	**Pro Pro**	**Phe**	**Pro**	**Trp**	Thr	Ser	Val
	Cys	Ser	**Leu**	**Leu**	**Leu**	**Pro Pro**	**Val**	**Pro**	**Trp**	Ala	Ile	Asp
	Leu	Pro	Tyr	**Leu**	Glu	**Pro Pro**	Asn	**Pro**	**Trp**	Ser	Gly	Cys
	Tyr	Trp	Arg	Ser	**Ala**	**Pro Pro**	**Leu**	**Pro**	**Tyr**	Phe	Phe	Leu
Consensus sequence												
	φ	Xxx	**Ψ**	**Ψ**	**Ψ**	**Pro Pro**	**Ψ**	**Pro**	**W/Y**	Xxx	Xxx	Xxx
VASP	**Ile**	His	**Leu**	Asp	**Trp**	**Pro Pro**	**Phe**	**Pro**	**Tyr**	Cys	**Tyr**	Val
	Thr	Ser	**Val**	Val	Gly	**Pro Pro**	Met	**Pro**	**Tyr**	Leu	**Phe**	His
	Leu	Cys	**Phe**	Gly	**Phe**	Ala	**Pro**	**Phe**	**Pro**	**Tyr**	Tyr	His	Ser
	Pro	Val	**Pro**	Thr	Met	**Pro Pro**	**Phe**	**Pro**	Leu	Asn	**Tyr**	Tyr
Consensus sequence												
	Ψ	Xxx	**Ψ**	Xxx	**Ψ/φ**	**Pro Pro**	**Phe**	**Pro**	**Tyr**	Xxx	**Y/F**	Xxx
Mena	Cys	Asn	Gly	**Phe**	**Leu**	**Pro Pro**	**Leu**	**Pro**	**Met**	**Tyr**	**Trp**	**Ala**
	Arg	Arg	Pro	**Leu**	**Leu**	**Pro Pro**	**Val**	**Pro**	**Met**	**Leu**	**Pro**	**Tyr**
	Thr	Ser	Gly	**Ala**	**Glu**	**Pro Pro**	**Leu**	**Pro**	**Met**	**Tyr**	**Trp**	**Ala**
	Gln	Leu	Phe	Gln	**Glu**	**Pro Pro**	Gly	**Pro**	**Met**	**Leu**	**Met**	**Tyr**
	Ile	Ser	Asn	**Leu**	**Asp**	**Pro Pro**	Cys	**Pro**	Asp	**Tyr**	**Val**	**Pro**
Consensus sequence												
	Xxx	Xxx	Xxx	**Ψ**	**α/L**	**Pro Pro**	**Ψ**	**Pro**	**Met**	**Y/L**	**Ψ**	**Ψ/Y**

Figure III-39: Sequences of single clones, after three rounds of standard panning on Evl, Mena and VASP EVH1 domains, and the deduced consensus sequences. The clones occur only once each. A residue is considered as representative for the consensus sequence, when it occurs in 50% or more of the sequences. Underlined amino acids are fixed residues within the CPLPPXP library. Ψ, α, and φ represent hydrophobic, acidic and aromatic amino acids, respectively. P_1 to P_{13} refer to the amino acid position within the sequence.

Even though only a small proportion of the sequenced clones comes from correctly formed variants, they show strong similarities specific to each target.

The variants isolated against Evl possess some determined positions, especially the position P_{10}, directly after the proline-core motif, which is highly defined as tryptophan or tyrosine. Except for this aromatic residue, Evl tends to have preferences for ligands being highly hydrophobic N terminal from the proline motif.

Similarly to Evl protein, VASP EVH1 domain shows a preference for an aromatic residue at the P_{10} position, where however, for the few clones analyzed, the residue is reduced as a tyrosine. Several other positions appear to be advantageous as an aromatic residue: a phenylalanine at P_8, between the prolines of the core motif, is highly representative of the VASP isolated clones, along with a tyrosine, or phenylalanine at position P_{12}. In addition, aromatic or hydrophobic residues might be profitable, at P_5, directly before the proline motif.

Mena exhibits slightly different preferences: the aromatic residue of P_{10} in Evl, and VASP, seems to be important for Mena, at P_{11} or P_{13}, as tyrosine. Each of the clones isolated against Mena, possess a tyrosine either at P_{11} or P_{13}. The P_{10} position is also highly defined, but as methionine. The sequence C terminal to the proline core appears to require tyrosine or hydrophobic residues to allow interaction with the Mena protein. The part N terminal of the prolines is less defined than for the other proteins, since only P_4 and P_5 show homologies along the aligned variants. P_4 is represented by hydrophobic residues, while P_5, is the only position where acidic residues are clearly enriched.

The overview of the consensus sequences of EVH1 domains, conceived from phage display analyses, puts in evidence some similarities. In addition, their alignment with the consensus sequences of the natural ligands of EVH1 domains involved in actin filament assembly, or in synaptic plasticity, allows some more observations:

Evl	**φ**	Xxx	**Ψ**	**Ψ**	**Ψ**	**Pro**	**Pro**	**Ψ**	**Pro**	**W/Y**	Xxx	Xxx	Xxx
VASP	**Ψ**	Xxx	**Ψ**	Xxx	**Ψ/φ**	**Pro**	**Pro**	**Phe**	**Pro**	**Tyr**	Xxx	**Y/F**	Xxx
Mena	Xxx	Xxx	Xxx	**Ψ**	**α/L**	**Pro**	**Pro**	**Ψ**	**Pro**	**Met**	**Y/L**	**Ψ**	**Ψ/Y**
Vesl	Xxx	**Arg**	**Ψ**	**Y/F**	**(Asp)**	**Pro**	**Pro**	**F/Ψ**	**Pro**	**W/Y**	Xxx	**Asp**	**Ψ**
EVH1 actin	*Xxx*	*Xxx*	***Ψ***	***α***	***Phe***	***Pro***	***Pro***	***Pro***	***Pro***	***Pro***	***Ψ***	***α***	***Glu/Ψ***
EVH1 synapse	*Xxx*	*Xxx*	*Xxx*	*Xxx*	*Xxx*	***Pro***	***Pro***	*Xxx*	***Pro***	***Phe***	***Arg***	***Asp***	

Very few acidic residues are present along the sequence of the isolated clones, compared to their proportion within the natural ligands (not reflected by the natural ligand consensus sequence, since the acidic residues are not restricted to one position). However, since none of

the clones has been dramatically enriched during the panning, this is not surprising.

The main difference between the natural ligand consensus sequences, and the one conceived after phage display, for Evl, Mena, VASP, and Vesl, lays in the aromatic residues. Within the natural ligands, a phenylalanine **before** the core motif is essential for interaction between EVH1 domains involved in actin filament assembly, and their ligands, while for Vesl EVH1 domain, the phenylalanine was located **after** the proline motif. Amino acid substitution studies show that phenylalanine before the proline motif is crucial for interaction between Mena and VASP with their ligands. Indeed, phenylalanine can only be replaced by other aromatic residues; tryptophan (which even improves the binding ability between both partners), and tyrosine. Every other substitution leads to a disruption of the interaction, except for Mena, which also allows a leucine at this position (Niebuhr et al., 1997).

Interestingly, the clones isolated with phage display, present, for all four EVH1 domains, an aromatic residue (preferentially tryptophan or tyrosine), however, **after** the proline motif. This was expected for Vesl, but this is the first time that such homologies are reported, between the two families of EVH1 domains.

However, when the clones were individually checked on phage ELISA for interaction with the different EVH1 domains, none of them show positive reaction with the EVH1 domains, which is not surprising, since none of the clones showed strong enrichment during affinity selection and phage ELISA is an assay only suitable for the ranking of clones having strong affinities.

2-2 Cosmix-plexing® recombination

Cosmix-plexing® recombination of the population selected after first and second rounds of panning, for each of the EVH1 domains, was performed. However, the screening of these optimized CPLPPXP library on EVH1 domains only allows the enrichment of the two non proline clones previously isolated against Nck1 and Nck2 SH3 domains:

NP1 Ser Thr Leu **Cys Asp Gly Tyr Cys**

NP2 Asn Pro Met Gly **Asp Gly Tyr** Phe

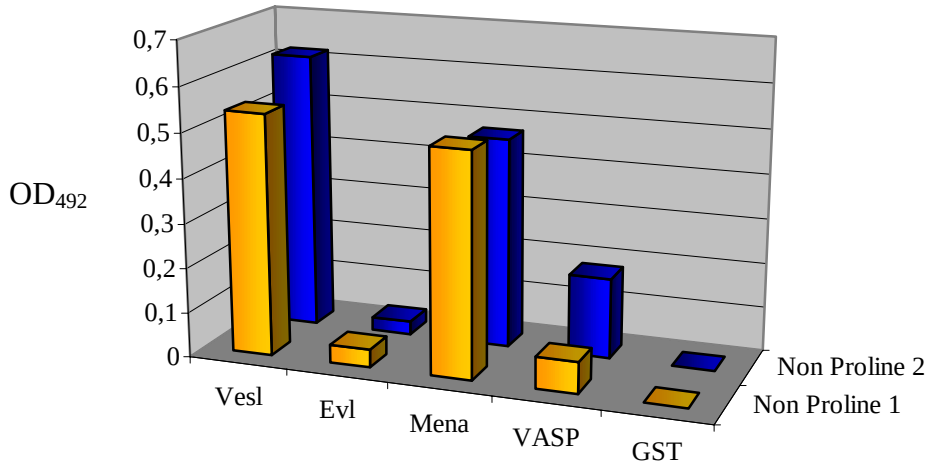
It has been thought that these two non proline clones might have been selected against EVH1 domains because the library was contaminated with these clones. Since there are smaller than the majority of the variants (only 8 amino acids, compared with 13), they have a selective advantage, due to a shorter replication time, and if no enrichment of interacting

partners occurs, they might represent a large proportion of the library, after a few rounds. It was therefore important to check for potential interaction between NP1, NP2, and the EVH1 domains.

2-3 Phage ELISA

NP1 and NP2 were individually propagated as single phage, and tested on phage ELISA against Evl, Mena, VASP, and Vesl EVH1 domains. The results of the analysis are summarized in Figure III-40.

Phage ELISA: affinity of non proline clones for EVH1 domains



	Phage ELISA of Non Proline clones against EVH1 domains	OD492 Vesl	Evl	Mena	VASP
NP1	Ser Thr Leu **Cys** **Asp** **Gly** **Tyr** **Cys**	0,54	0,04	0,50	0,07
NP2	Asn Pro Met Gly **Asp** **Gly** **Tyr** Phe	0,62	0,03	0,47	0,18
Consensus sequence	Xxx Xxx Xxx Xxx **Asp** **Gly** **Tyr** Xxx				

Figure III-40: Relative affinity of the non proline clones with EVH1 domains, according to phage ELISA. GST-EVH1 proteins were coated in the wells. An equivalent amount of phage were incubated in every well (10^{10} phage/well). After several washes, the remained bound phage were detected with anti-M13 antibody. The absorbance was detected at 492 nm. A well coated with GST alone was used as a control, to ensure that the single clones bind to the EVH1 domains, and not to the GST. The OD values are the average of triplicates.

Surprisingly, the non proline clones react positively with the EVH1 domains, even though very weak with Evl EVH1 domain. These results bring new interest to the field of proline rich ligands binding proteins, since it is the first time that a link between SH3 and EVH1 domains can be established. Until now, the only common point between both kind of domains, was that they both interact with proline rich ligands. However, these ligands showed different consensus sequences. They were considered as two distinct families, since Niebuhr (1997) showed that SH3 ligands do not interact with EVH1 domains. But clones were isolated which are able to bind with both SH3 and EVH1 domains, and furthermore, which are not proline rich!

The enrichment of the non proline (NP) clones, over proline rich variants, during the screening of the optimized CPLPPXP libraries, tends to indicate that the NP clones showed higher affinity for the EVH1 domains than other clones present within the library, and therefore, might explain the apparent failure of the panning selection procedure.

2-4 Vesl cosmix plexing library

Since the Evl, VASP, and to a smaller extent, Mena proteins, show consensus sequences similar to the one obtained for Vesl (regarding the position of aromatic residue after the proline core motif), it was decided to screen the cosmix-plexed library optimized for Vesl EVH1 protein, against Evl, VASP, and Mena.

The use of this library allows the isolation of new clones for Evl, and VASP, some of which occurred several times. However, curiously, the output and input values did not show any enrichment. The panning of this Vesl library against Mena was not successful. Most of the clones were unproductive clones, and the other one did not show any homology among themselves.

The clones isolated from Vesl cosmix-plexed library against Evl, show interesting similarities, which lead to the conception of two consensus sequences, as presented in Figure III-41. The clones identified against Evl, are very different from the trend obtained after normal panning. They were classified into two groups, according to their sequences.

The first group gathers the clones which exhibit a serine after the proline core motif, at P_{10}, whereas the second group comprises the variants showing an aspartic acid at P_{10}.

Except the P_{10} position, the first group shows an aromatic residue at P_{12}, and Cys at P_8, between the prolines. However, both of these positions might be not significant, since, out of the three clones which form this group, the dominant clone (which occurs 7 times, compared

to one for both other clones) does not contain any either of these P_8 and P_{12} preferences shown by the two other clones.

The second group appears to be slightly more defined than the previous one. Except the Asp at P_{10}, a phenylalanine at P_8 is clearly determined, and hydrophobic residues at $P_{3, 4, 11,}$ and $_{12}$, seem to be advantageous.

Sequences against Evl EVH1 domain, panning with Vesl cosmix-plexing® library												Frequency
P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}	P_{13}	
First group													
Phe	Cys	Asp	Ile	Gly	**Pro**	**Pro**	Val	**Pro**	**Ser**	Asp	**Val**	Pro	7x
Cys	Tyr	Ser	Ile	Glu	**Pro**	**Pro**	**Cys**	**Pro**	**Ser**	Ser	**Tyr**	Tyr	1x
Arg	Ala	His	Gln	Trp	**Pro**	**Pro**	**Cys**	**Pro**	**Ser**	Leu	**Phe**	Cys	1x
Consensus sequence													
Xxx	Xxx	Xxx	**Ile**	Xxx	**Pro**	**Pro**	**Cys**	**Pro**	**Ser**	Xxx	**Ψ/φ**	Xxx	
Second group													
Gly	Cys	**Pro**	**Phe**	Ser	**Pro**	**Pro**	**Phe**	**Pro**	**Asp**	**Val**	**Ala**	Gly	3x
Leu	Val	Thr	Gly	Arg	**Pro**	**Pro**	**Phe**	**Pro**	**Asp**	His	Ser	Arg	2x
Tyr	Gln	**Ala**	**Ala**	Ala	**Pro**	**Pro**	**Phe**	**Pro**	**Asp**	**Val**	**Ala**	Gly	1x
Ser	His	**Ile**	**Phe**	Asp	**Pro**	**Pro**	**Phe**	**Pro**	Tyr	Gly	**Pro**	Met	1x
Consensus sequence													
Xxx	Xxx	**Ψ**	**Ψ**	Xxx	**Pro**	**Pro**	**Phe**	**Pro**	**Asp**	**Ψ**	**Ψ**	Xxx	

Figure III-41: Sequences of single clones, after three rounds of panning on Evl EVH1 domain, with the Vesl cosmix-plexed library, and the deduced consensus sequences. The frequency of the clones are indicated. A residue is considered as representative for the consensus sequence, when it occurs in 50% or more of the sequences. Underlined amino acids are fixed residues within the CPLPPXP library. Ψ, and φ represent hydrophobic and aromatic amino acids, respectively. P_1 to P_{13} refer to the amino acid position within the sequence.

The clones isolated from Vesl cosmix-plexed library against VASP are the most interesting one. Here also, they lead to the conception of two consensus sequences (Figure III-42), but the tendency observed after the standard panning, is detected here as well, giving more credibility to the derived consensus. The consensus sequence of VASP after normal panning, is repeated here, to help the comparison.

Ψ Xxx **Ψ** Xxx **Ψ/φ** **Pro** **Pro** **Phe** **Pro** **Tyr** Xxx **Y/F** Xxx

From both groups, the P_{10} position is clearly determined as tryptophan, or tyrosine, as it was after standard panning. However, the Vesl cosmix-plexed library allows the proposal of long consensus sequences. In addition, more acidic residues are found all along the sequence. The two groups were distinguished according to their preferences within the region N terminal from the proline rich motif.

The first group shows a clear consensus sequence as **Asp Arg Leu Tyr Asp/Glu** before the proline motif. It has to be noticed that two acidic residues are necessary within this short region, probably to stabilize the interaction between the ligands and the VASP domain. Interestingly, a tyrosine is found at position P_4, which might build a link between the phage display isolated clones, and the natural ligands, which all possess a phenylalanine directly before the proline motif, but which can be replaced by tryptophan or tyrosine (Niebuhr et al., 1997).

	Sequences against VASP EVH1 domain, panning with Vesl cosmix plexing library													**Frequency**
	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}	P_{13}	
First group														
	Ser	Leu	**Leu**	**Tyr**	**Glu**	**Pro**	**Pro**	**Pro**	**Pro**	**Trp**	Asn	Ser	**Pro**	1x
	Phe	**Arg**	Gly	Gln	**Glu**	**Pro**	**Pro**	**Phe**	**Pro**	**Trp**	Leu	Pro	**Trp**	1x
	Asp	**Arg**	**Leu**	**Tyr**	**Asp**	**Pro**	**Pro**	**Ala**	**Pro**	**Tyr**	Arg	Ile	Cys	1x
	Asp	**Arg**	**Leu**	**Tyr**	**Asp**	**Pro**	**Pro**	**Phe**	**Pro**	**Tyr**	Gly	Pro	**Met**	1x
	Asp	**Arg**	His	**Tyr**	Arg	**Pro**	**Pro**	**Phe**	**Pro**	**Trp**	Ala	Asp	Gly	* 2x
Consensus sequence														
	Asp	**Arg**	**Leu**	**Tyr**	**α**	**Pro**	**Pro**	**F/Ψ**	**Pro**	**W/Y**	Xxx	Xxx	**Ψ**	
Second group														
	Val	**Leu**	**His**	**Val**	**Arg**	**Pro**	**Pro**	Thr	**Pro**	**Trp**	**Val**	Ser	Val	3x
	Asp	Arg	**His**	**Tyr**	**Arg**	**Pro**	**Pro**	Phe	**Pro**	**Trp**	**Ala**	Asp	Gly	* 2x
	Leu	**Ile**	**Lys**	**Val**	**Arg**	**Pro**	**Pro**	Val	**Pro**	**Trp**	**Leu**	Phe	Cys	1x
	Asn	**Leu**	**Arg**	**Tyr**	**Arg**	**Pro**	**Pro**	Ser	**Pro**	Asp	Arg	Tyr	Ser	1x
Consensus sequence														
	Xxx	**Leu**	**β**	**Y/V**	**Arg**	**Pro**	**Pro**	Xxx	**Pro**	**Trp**	**Ψ**	Xxx	Xxx	

Figure III-42: Sequences of single clones, after three rounds of panning on VASP EVH1 domain, with the Vesl cosmix-plexed library, and the deduced consensus sequences. The frequency of the clones are indicated. A residue is considered as representative for the consensus sequence, when it occurs in 50% or more of the sequences. Underlined amino acids are fixed residues within the CPLPPXP library. Ψ, α, and β represent hydrophobic, acidic, and basic amino acids, respectively. P_1 to P_{13} refer to the amino acid position within the sequence. The clone marked with a red asterisk, possess a sequence which belong to both groups.

The clones of the second group also allow the conception of a clear consensus sequence. But, whereas in the first group, the proline region N terminal sequence was determined with 2 acidic residues, and one basic (P_1 and P_5 for the acidic residue, and P_2 for the arginine), for the second group, the basic residues appear to be advantageous, since two positions, P_3 and P_5 are defined as basic, as observed on the consensus sequence:

Leu β Tyr/Val Arg Pro Pro Xxx Pro Trp Ψ

In both groups, the importance of tryptophan or tyrosine at P_{10} is demonstrated, but they also both show a high tendency of exhibiting tyrosine at P_4. Both groups have a common consensus sequence, resumed with **Tyr Xxx Pro Pro Xxx Pro Trp/Tyr.** The clone marked with an asterisk is very interesting. Indeed, its sequence possesses characteristic belonging to both groups. The Asp_1 and Arg_2 are characteristics from the first group, while His_3 and Arg_5 belong to the second group. Tyr_4, the proline core motif, and Trp_{10} were already defined as advantageous for both set of clones.

Unfortunately, when the clones were tested for relative affinity with Evl, and VASP EVH1 domains, on phage ELISA, the clones were not reacting. This negative reaction suggests that the clones enriched during the panning selection are weak binders, since the phage ELISA is an assay only relevant for high affinity ligands. However, this low affinity binding remains unclear, since the dominant clones might have been considered as preferential binding partners for the interaction with the concerned target.

F- Characterization of two non proline clones: NP1 and NP2

Two unexpected clones have been, in several cases, isolated against SH3 and EVH1 domains. The oddness, and particularity of these clones, is that they are not proline rich, as SH3 and EVH1 domain ligands are assumed to be.

These clones, NP1 and NP2, have been previously reported in this work, since they have been enriched during CPLPPXP library screening, against Nck1 and Nck2 SH3 domains, but also Evl, Mena, and VASP EVH1 domains. However, NP1 and NP2 were isolated more frequently. Panning selections have been performed against other SH3 domains, like Abl,

p67phox, and ALP. For these proteins, the non proline clones were the only identified enriched variants.

It has already been shown by phage ELISA, that NP1 and NP2 specifically bind to Nck1, Nck2, and the various EVH1 domains (Figures III-26, and -40).

Phage ELISA analyses have been extended, in order to further characterize NP1 and NP2 clones.

1- Phage ELISA

NP1 and NP2 were individually propagated as single phage, and tested on phage ELISA, against various SH3 and EVH1 domains, to check which proteins are able to interact with these non proline clones, and to compare their relative affinity (Figure III-43).

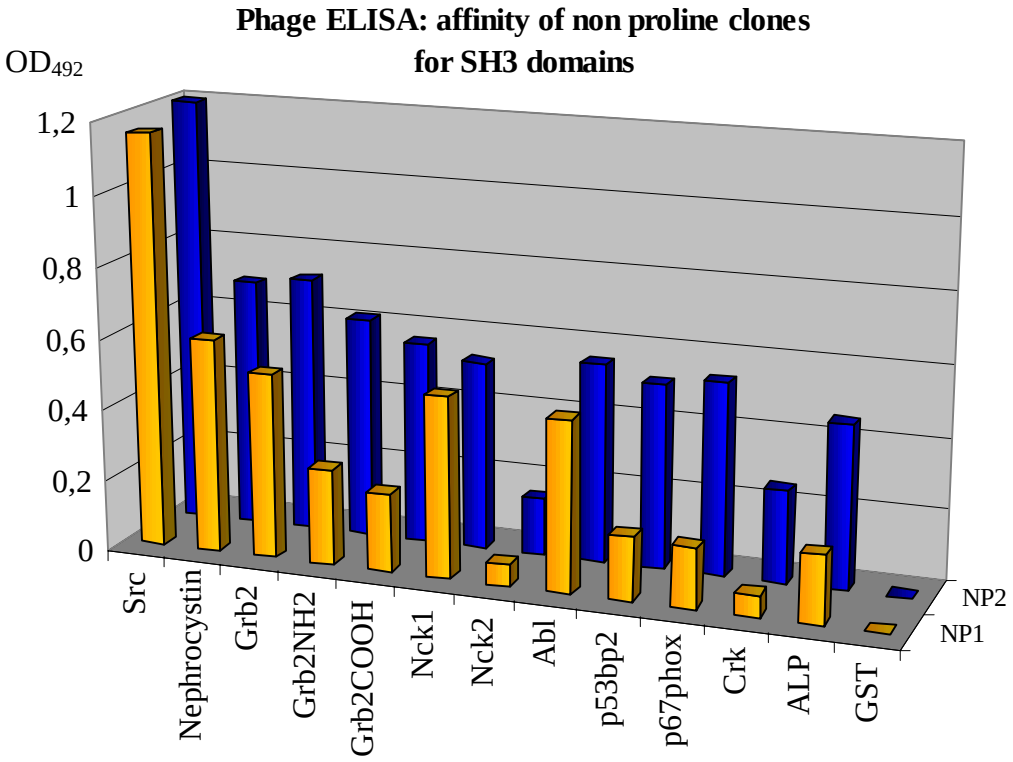
Both non proline clones specifically interact with SH3 and EVH1 domains, since the GST alone is not able to bind with NP1 or NP2. In most of the cases, NP2 shows slightly better binding capacities than NP1.

The various SH3 and EVH1 domains react differently with the non proline clones. Src SH3 domain is the protein which shows the strongest interaction with the non proline clones. In the contrary, Nck2 and Crk SH3 domains, and VASP, but particularly Evl EVH1 domains react very weakly with the non proline clones.

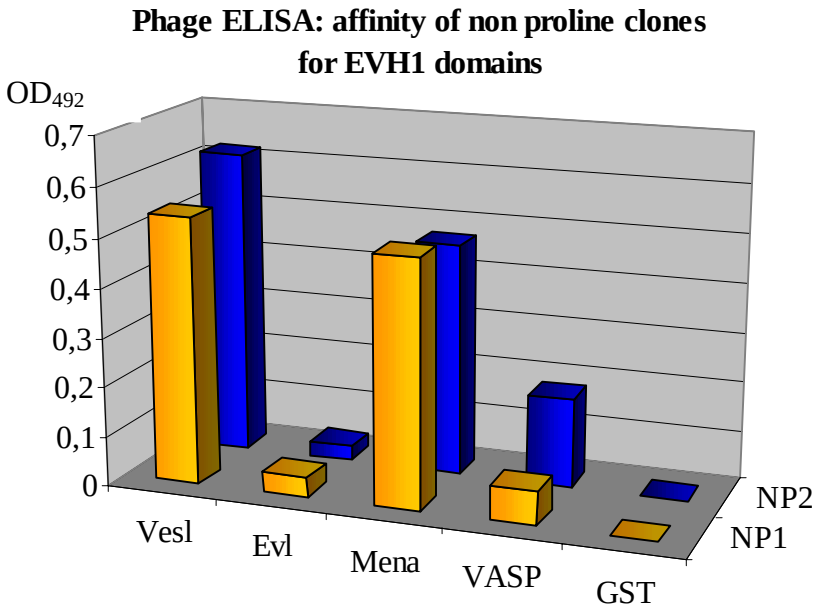
The phage ELISA analyses, demonstrate that NP1 and NP2 are really able to interact with SH3 and EVH1 domains. However, this gives no information on their site of interaction. The following analyses were performed in an attempt to answer this question.

Figure III-43: Relative binding of the non proline clones for SH3 (A), and EVH1 (B) domains, according to phage ELISA. GST-SH3, -EVH1 proteins were coated in the wells. An equivalent amount of phage were incubated in every well (10^{10} phage/well). After several washes, the remained bound phage were detected with anti-M13 antibody. The absorbance was detected at 492 nm. A well coated with GST alone was used as a control, to ensure that the single clones bind to the SH3 or EVH1 domains, and not to the GST. The OD values are the average of triplicates. (C-) The sequences of both non proline clones, and deduced consensus sequence, are shown.

A



B



C

Non Proline clones								
NP1	Ser	Thr	Leu	**Cys**	**Asp**	**Gly**	**Tyr**	**Cys**
NP2	Asn	Pro	Met	Gly	**Asp**	**Gly**	**Tyr**	Phe
Consensus sequence	Xxx	Xxx	Xxx	Xxx	**Asp**	**Gly**	**Tyr**	Xxx

2- Competitive phage ELISA

In this type of phage ELISA, the interaction between a single clone and the protein of interest is analyzed, in presence of increasing amounts of competitor. The S1 clone, which was the clone showing the strongest binding capacities for Src SH3 domain, in phage ELISA test (Figure III-23), as been synthesized as free peptide, and is used as competitor in this test.

The Src SH3 domain has been immobilized on a microtiter plate, and incubated with a fixed concentration of the phage variants S1, S2, S3, and S4. After washing, the S1 free peptide was added to the phage particles attached to the Src SH3 domain. Various concentrations of S1 free peptide were incubated with the phage/protein complex (0 to 1000 µM). The amount of phage which remain bound, was then evaluated, after new washing steps, by using an anti-M13 antibody.

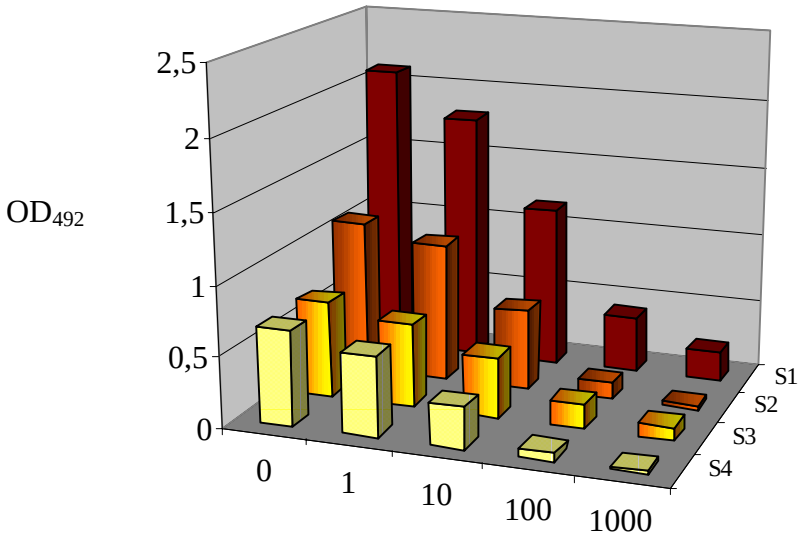
Figure III-44 shows the evolution of the binding disruption between the Src protein, and different phage ligands, as a function of the amount of competitor. As the concentration of S1 synthetic peptide increases, the amount of attached phage decreases. 10 µM of free peptide added to the wells, is enough to compete around 50% of the phage, away from the Src binding site (Figure III-44/B). This is shown for the four clones (S1 to S4) which were able to positively react with the Src SH3 domain in phage ELISA, demonstrating that all four clones bind to the same site of Src, as expected, in view of the very high similarities between their sequences.

These results also demonstrate that the synthetic S1 peptide exhibits the same binding properties as the linear peptide displayed on the surface of the phage. This is not always true, since in some cases, the linear displayed peptide is constrained by its local environment.

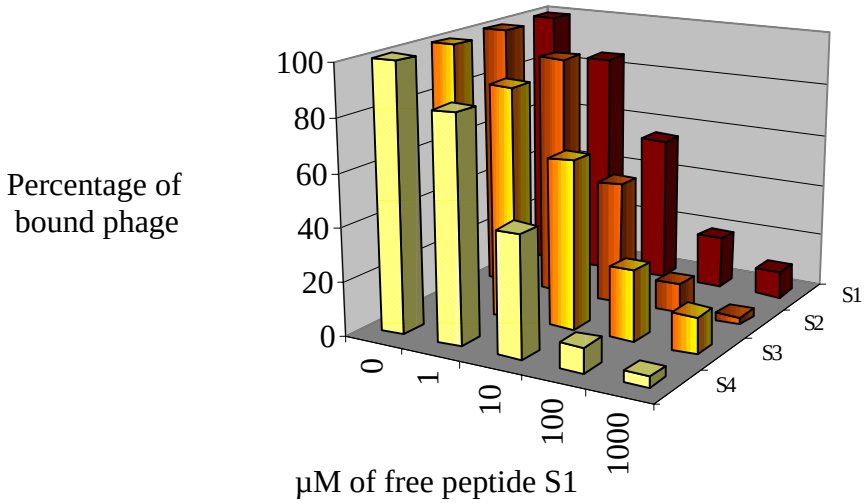
The same experiment was repeated, to investigate if the non proline clone attachment to Src SH3 domain is disrupted by the S1 synthetic peptide. Src was immobilized, and incubated with both non proline clones, NP1 and NP2. After washing, various concentrations of the competitor were added, and the amount of bound phage remaining, evaluated. The results of the analysis are presented in Figure III-45.

Figure III-44: Competitive phage ELISA of Src positive clones, on Src SH3 protein. GST-Src SH3 domain has been immobilized on a microtiter plate, and incubated with the Src positive phage particles. An equivalent amount of phage were incubated in every well (10^{10} phage/well). After washings, S1 synthetic peptide was added to the complex phage/Src SH3 domain, at various concentrations (0 to 1000 µM). The amount of remained bound phage was then detected, after new washing steps, with anti-M13 antibody. The absorbance was detected at 492 nm. The OD values are the average of triplicates. (A-) Relative affinity of S1 to S4 clones with Src SH3 domain, according to the concentration of competitor. (B-) Percentage of remained bound phage (S1 to S4) to Src SH3 domain, according to the concentration of S1 synthetic peptide. (C-) The sequences of clones and synthetic peptide, are given.

A **Competitive phage ELISA with free peptide S1**



B **Competitive phage ELISA with free peptide S1**



.

C **Phage interacting with Src SH3 domain**

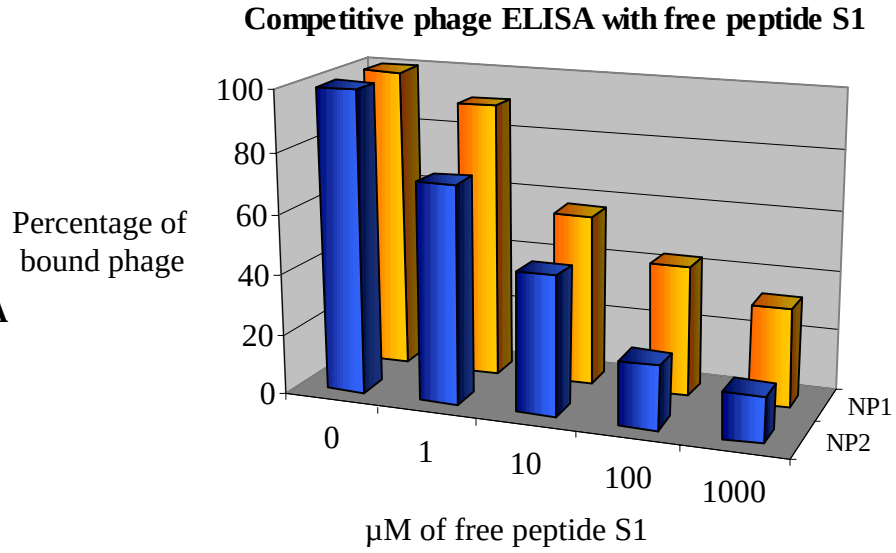
S1		Gly	Asn	**Arg**	**Pro**	**Leu**	**Pro**	**Pro**	**Ile**	**Pro**	Ser	His	**Pro**	**Phe**		
S4		Lys	Gly	**Arg**	**Pro**	**Leu**	**Pro**	**Pro**	**Val**	**Pro**	Gly	Thr	**Pro**	**Ser**		
S2				**Arg**	**Pro**	**Leu**	**Pro**	**Leu**	**Pro**	**Pro**	**Val**	**Pro**	Trp	**Val**	Arg	Trp
S3				**Arg**	**Ser**	**Leu**	**Pro**	**Leu**	**Pro**	**Pro**	**Val**	**Pro**	Thr	**Ser**	Ser	Ala

S1 synthetic peptide - Competitor

Gly	Asn	**Arg**	**Pro**	**Leu**	**Pro**	**Pro**	**Ile**	**Pro**	Ser	His	**Pro**	**Phe**

As can be seen in Figure III-45, the synthetic peptide disrupts the interaction between Src SH3 domain, and its ligands, NP1 and NP2. This clearly demonstrates that the non proline clones interact with Src domain at the same binding site as the proline rich ligands. However, even with a very high concentration of competitor (1000 µM), the non proline clones are not completely competed. Approximately 15 and 30% of the phage remain attached to Src (for NP2 and NP1 respectively), whereas competition of S1 to S4 clones, with 1000 µM of S1 free peptide, allows only approximately 3 to 14% of the S1 to S4 phage to remain bound to the Src SH3 domain. In addition, since the relative binding capacities of NP1 and NP2 for Src are weaker than that of the S1 clone, a more complete competition was expected.

A



B

Non Proline clone sequences								
NP1	Ser	Thr	Leu	**Cys**	**Asp**	**Gly**	**Tyr**	**Cys**
NP2	Asn	Pro	Met	Gly	**Asp**	**Gly**	**Tyr**	Phe

S1 synthetic peptide - Competitor												
Gly	Asn	**Arg**	**Pro**	**Leu**	**Pro**	**Pro**	**Ile**	**Pro**	Ser	His	**Pro**	**Phe**

Figure III-45: Competitive phage ELISA of non proline clones, on Src SH3 protein. GST-Src SH3 domain has been immobilized on a microtiter plate, and incubated with the Non Proline clones. An equivalent amount of phage were incubated in every well (10^{10} phage/well). After washings, S1 synthetic peptide was added to the complex phage/Src SH3 domain, at various concentrations (0 to 1000 µM). The amount of remained bound phage was then detected, after new washing steps, with anti-M13 antibody. The absorbance was detected at 492 nm. The OD values are the average of triplicates. (A-) Percentage of remained bound phage (NP1 and NP2) to Src SH3 domain, according to the concentration of S1 synthetic peptide. (B-) The sequences of clones and synthetic peptide, are given.

However, one can assume that the non proline clones bind each of the SH3 domains tested, by contacting the binding site known to interact with the proline rich ligands. Similarly, they might also bind EVH1 domains also at their active binding site, but this was not tested.

It also has to be emphasized, that the non proline clones bind different SH3 and EVH1 domains with variable relative affinity (Figure III-43). This suggests that even if these clones bind in general to each of the SH3 and EVH1 domains, they might however interact with them in a protein specific way, meaning that they might not solely bind to the grooves which make contact with the proline core motif, but that they involve in the interaction, all or a part of the specificity pocket (Figure I-10).

G - BIAcore analysis – real time interaction measurements

To evaluate the binding properties of the phage displayed ligands, the relative binding of phage isolates was measured by BIAcore. As previously shown by Rickles et al. (1994; 1995), SH3 binding can be studied when the ligands are displayed on the surface of the bacteriophage. The strategy chosen to investigate the SH3/phage-ligand interactions, is illustrated in Figure III-46. The phage displayed ligands which had been shown to interact positively to SH3 domains (Src and Nephrocystin SH3 domains), on phage ELISA (though weakly), were tested on BIAcore. Since Src and nephrocystin are both immobilized on different channels of the chip, evaluation of the cross reaction is directly feasible. Negative controls were performed, to ensure that the phage ligands bind the SH3 domains, and not the GST, or the anti-GST antibody used to trap the GST-SH3 fusion proteins.

The Figure III-47 is a summary of the BIAcore analysis performed with this strategy. The phage displayed ligands which interact with Src or nephrocystin SH3 domains under the BIAcore conditions, are presented in this figure, and it is then possible to compare their relative affinity for their target protein.

It has to be noticed that not all clones which were positive under phage ELISA conditions, positively react in the BIAcore analysis. Only the strongest ELISA ligands show a high affinity for the SH3 domains on real time interaction measurements. However, the classification of the ligands, according to their relative binding on phage ELISA (Figures III-31, -32), corresponds in general to the BIAcore classification.

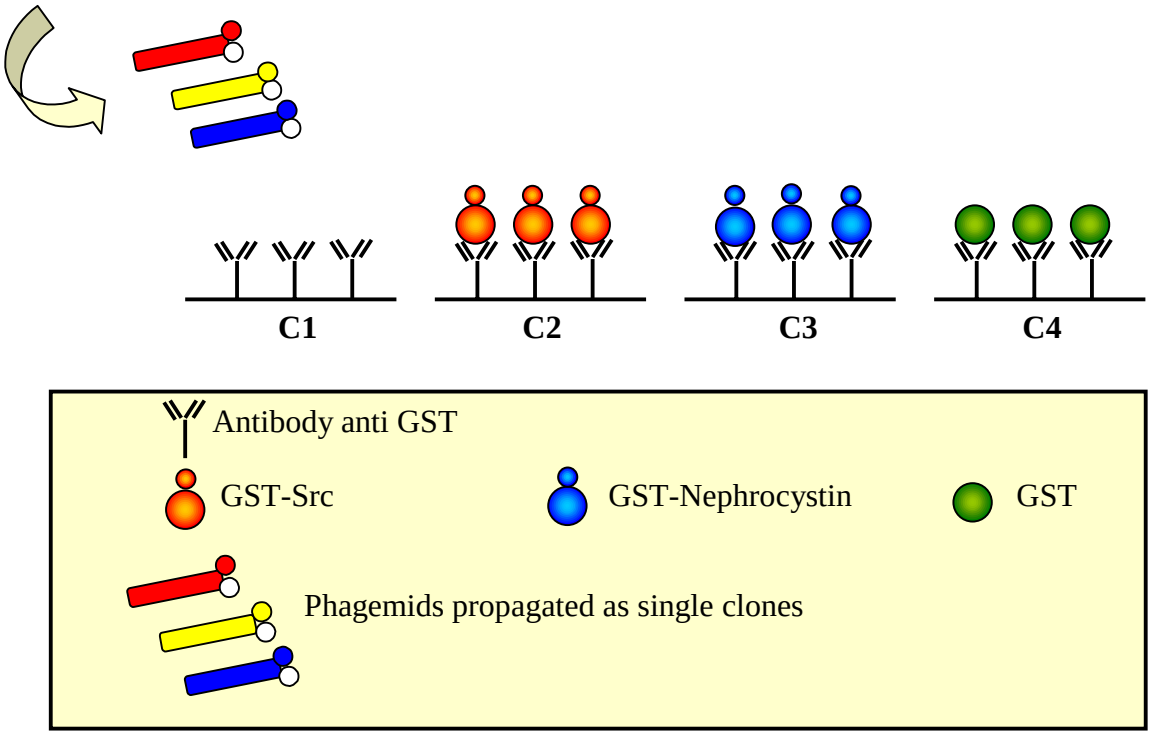
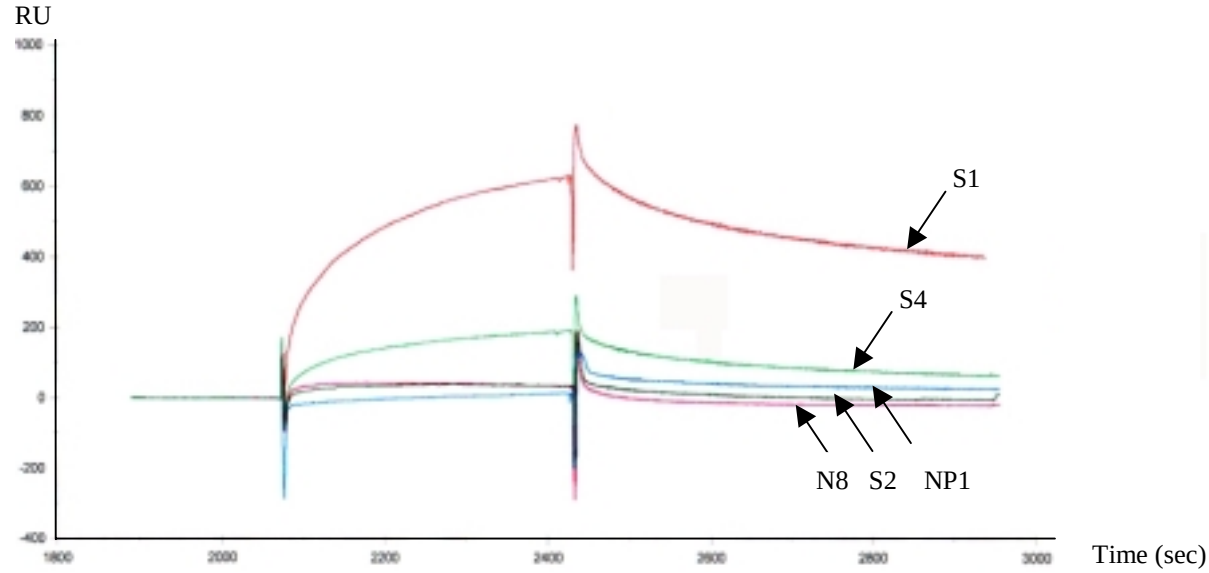


Figure III-46: Illustration of the strategy used to determine which phage displayed peptides have a high affinity for Src and nephrocystin SH3 domains, according to BIAcore measurements. A F1 biosensor chip was coated on its four channels (C1 to C4), with antibody anti GST. The channels C1 and C4 are reference channels. On C1, one can ensure that the phage ligands do not bind to the anti-GST antibody, while on C4, a potential interaction between the phage ligands and the GST is investigated. On C2 and C3, the GST-Src and -Nephrocystin SH3 proteins are immobilized on the anti-GST antibody. Then, a determined concentration of a phage ligand is added to the medium which passes through the four channels of the chip. The potential interactions with the immobilized proteins, are monitored and evaluated, due to the altered beam angle (plasmon resonance effect). The phage and proteins are then washed away with a glycine buffer (only the antibodies remain attached to the chip), and a new cycle is performed: new proteins are immobilized, and another phage displayed ligand is given.

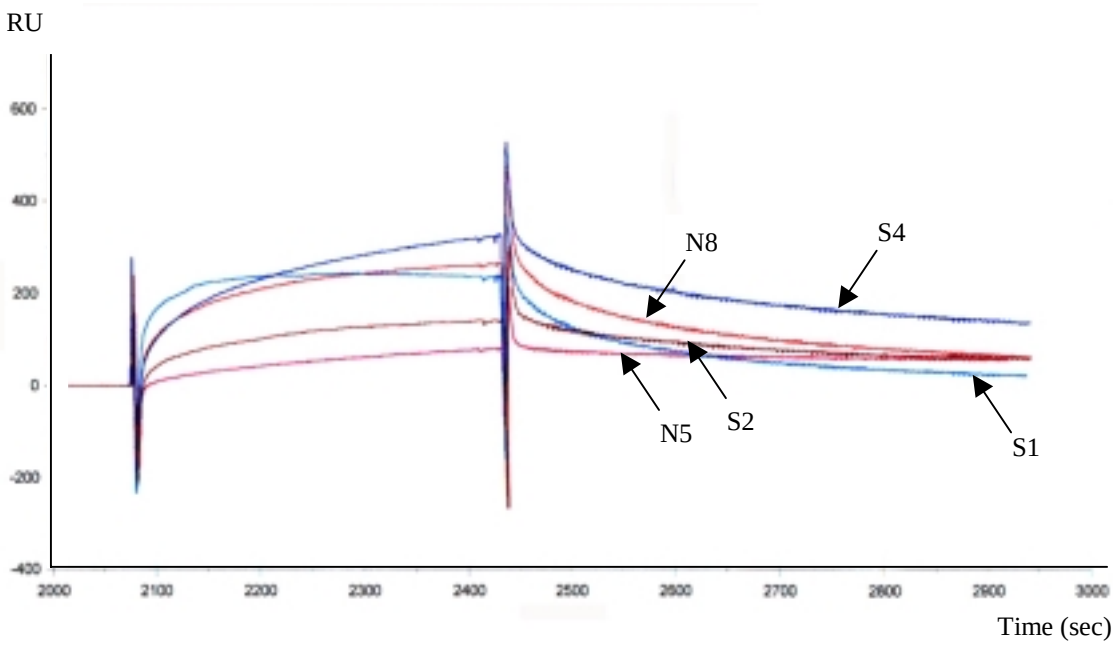
The S1, S2, S4, N8, and the Non Proline 1 (NP1) phage displayed peptides, interact with Src SH3 domain, on BIAcore measurements (Figure III-47), the ligands with the higher relative affinity, being S1, and then S4. This leads to the conception of a very clear consensus sequence: **Arg Pro Leu Pro Pro Ψ Pro** Xxx Xxx **Pro**, where Ψ represents the hydrophobic residues, and the underlined prolines, were fixed with the CPLPPXP library.

Figure III-47: Binding of phage displayed peptides to Src and nephrocystin SH3 domains, measured by BIAcore. Interaction between the immobilized SH3 domain, and the ligand (phage displayed peptide), is indicated by an increase of the RU value (Resonance Units). (A-) Relative affinity of positive clones for Src SH3 domain. (B-) Relative affinity of positive clones for nephrocystin SH3 domain. (C-) Sequences of the positive clones are given.

A **Immobilized Src SH3 domain – positive phage displayed peptides**



B **Immobilized Nephrocystin SH3 domain – positive phage displayed peptides**



C **Positive clones for Src and/or Nephrocystin SH3 domain, with BIAcore analysis**

S1	Gly	Asn	**Arg**	**Pro**	**Leu**	**Pro**	**Pro**	**Ile**	**Pro**	Ser	His	**Pro**	**Phe**			
S4	Lys	Gly	**Arg**	**Pro**	**Leu**	**Pro**	**Pro**	**Val**	**Pro**	Gly	Thr	**Pro**	Ser			
S2			**Arg**	**Pro**	**Leu**	**Pro**	**Leu**	**Pro**	**Pro**	**Val**	Pro	Trp	**Val**	Arg	Trp	
N8			***Arg***	**Pro**	**Leu**	**Pro**	**Pro**	Thr	**Pro**	**Pro**	Ile	**Pro**	Gly	Ser	Gly	Ser
N5			**Arg**	**Met**	**Leu**	**Pro**	**Val**	Tyr	**Pro**	**Pro**						
NP1			*Ser*	*Thr*	*Leu*	***Cys***	***Asp***	***Gly***	***Tyr***	***Cys***						

Several clones interact with the nephrocystin SH3 domain. Every ligand which was able to bind to Src under BIAcore conditions (except NP1), was able to interact with nephrocystin. The N5 phage interacts with nephrocystin. It is the only clone tested which did not show cross-reaction between Src and nephrocystin SH3 domains. It is assumed that the reason why N5 is unable to react on BIAcore to Src SH3 domain, is that the absence of Pro between the Arg, and the Leu, just before the proline rich core motif, disrupts a high affinity interaction, and therefore, the binding appears to be too weak to be detected on BIAcore measurements.

The two strongest ligands for nephrocystin, are S4 and N8, leading to a consensus sequence highly homologous to that of Src: **Arg Pro Leu Pro Pro** Xxx **Pro** Xxx Xxx **Pro**.

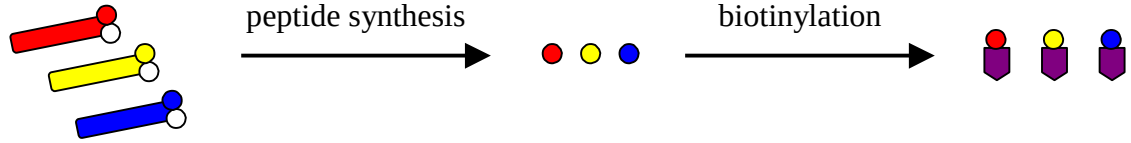
Interestingly, the S1 ligand reacts differently to the other clones. The association between nephrocystin and S1 is very fast, compared to the other clones, and a complete saturation of the immobilized nephrocystin seems to be reached. In addition, the dissociation rate is also faster than for the other ligands.

This initial BIAcore analysis allows the distinction of the phage displayed peptides which interact strongly with the Src or the nephrocystin SH3 domains. However, this investigation cannot be used to evaluate the affinity constants of the different clones for their targets. In theory, a monovalent expression of the peptide, on the surface of the bacteriophage is reached, during the particle assembly. However, it cannot be excluded that some particles display no peptide, and others, two or even more of them. In this latter case, an avidity phenomenon might amplify the observed signal, and a false interpretation of the affinity constants might result. To avoid this problem, a second BIAcore strategy has been developed, which takes advantage of the first selected ligands (Figure III-48). In this technique, to ensure an interaction between a single peptide, and the target SH3 domain, synthetic peptides are prepared, corresponding to sequences of the positive selected clones.

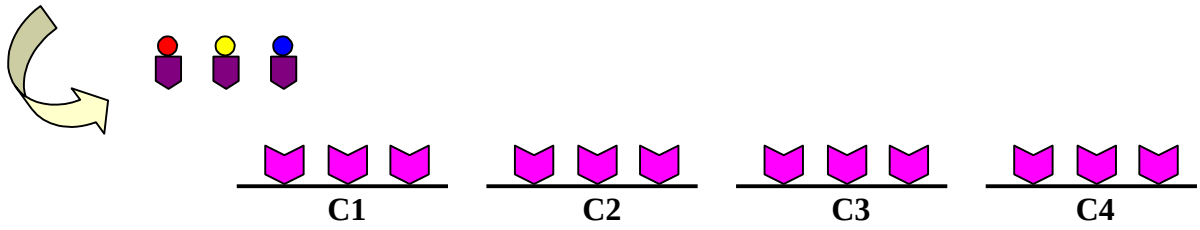
In BIAcore measurements, the interaction between the partners, is detected by an alteration of the laser beam when adsorption of biomolecules on the sensor chip takes place. Such changes are directly proportional to the amount and molecular weight of the macromolecules bound at any time. For this reason, it has been chosen to immobilize the peptides, and to add the SH3 target proteins in the constant flow medium, since the larger size of the GST fusion proteins will give a stronger signal, and help the determination of the affinity constants. A very small amount of peptide was immobilized (around 20 RU), in order to easily reach an association saturation between both partners, since this phase is necessary to calculate the constants. To facilitate the coating, the peptides were biotinylated, and

immobilized on a streptavidin sensor chip.

A- Preparation of synthetic peptides, from the phage displayed peptides



B- Immobilization of the biotinylated peptides, on streptavidin chip



C- Affinity test with the proteins: GST-Src, GST-Nephrocystin, GST

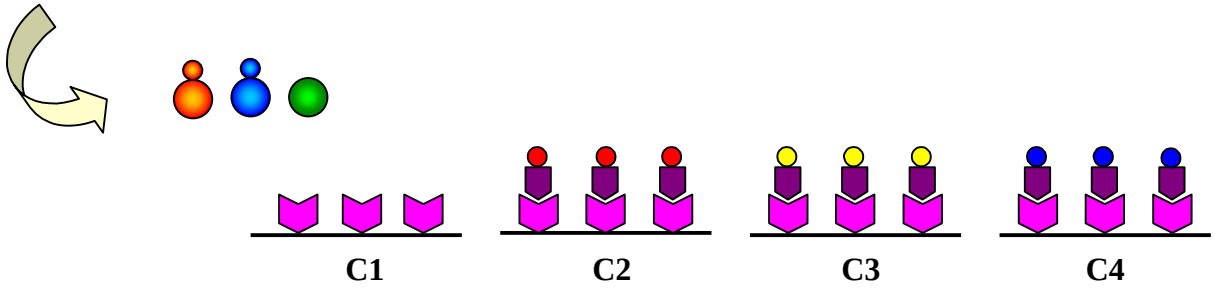


Figure III-48: Illustration of the strategy used to determine the affinity constants of Src and nephrocystin peptides, according to BIAcore measurements. (A-) Previously selected phage displayed peptides are synthesized as free peptide, and conjugated to biotin. (B-) The biotinylated peptides are immobilized on a streptavidin biosensor chip. The first channel (C1) is kept without peptide, as reference. Different biotinylated peptides are immobilized on C2 to C4. (C-) A determined concentration of GST-Src, GST-Nephrocystin, or GST alone (as negative control) is added to the medium which go through the four channels of the chip. The potential interactions with the immobilized proteins, are monitored and evaluated, due to the altered beam angle. The phage and proteins are then washed away with a HCl buffer.

Then, a panel of different concentrations of SH3 target proteins are distributed within the medium which goes through the four channels of the streptavidin sensor chip, and the resulting relative binding is compared, in order to determine the affinity constants. The Figure III-49 illustrates the binding of Src SH3 domain to the synthetic biotinylated peptide S1, according to its concentration.

On each of the curve of the sensorgram showed in Figure III-49, the saturation response values (Req) are obtained, from report points in the steady state region (when the association equilibrium has been reached), for each of the Src concentration tested. A plot of Req versus

the concentration of the injected proteins (Src or nephrocystin) is created, and used to estimate the K_D of the complex. The Rmax, which is the maximum response expected when all ligands on the surface are retaining an injected ligand molecule, can be deduced from the equilibrium phase of the steady state affinity data curve. When the Rmax data, divided by 2, is reported on Req axis of the diagram, it cuts the curve on a particular point, from which the coordinate on the concentration axis, give a direct, and very precise evaluation of the K_D value (as shown in Figure III-50/B).

Immobilized biotinylated S1 peptide – relative affinity of Src SH3 domain

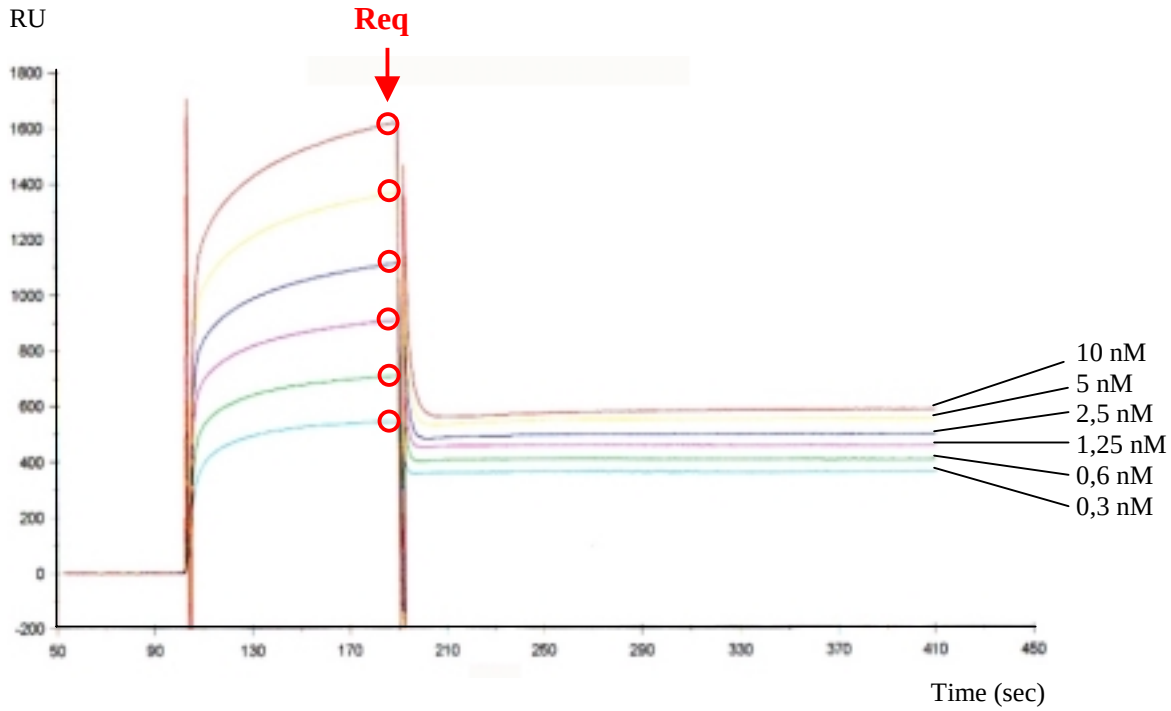


Figure III-49: Binding of Src SH3 domain, to immobilized synthetic peptide S1, measured by BIAcore. Interaction between the immobilized S1 peptide, and the Src SH3 domain, is indicated by an increase of the RU value (Resonance Units). Concentrations of Src SH3 domain, varying from 0,3 to 10 nM, were injected to the coated S1 peptide. The response at equilibrium (Req) is estimated, for each of the Src concentration, and will be used for a steady state affinity data evaluation.

As shown in Figure III-50/A, the curve resulting from the Req values determined in Figure III-49, is not satisfying, since the Src concentrations used, were too weak to reach the equilibrium phase. The experiment was repeated with Src concentrations reaching 100 nM. However, this was possible only once, since our Src stock was not large enough. It was then chosen to analyze the interaction between the S1 peptide, and the Src SH3 domain, since S1 appears to be the best interacting ligand that we isolated against Src. The deduced Req versus Src concentration curve is presented in Figure III-50/B.

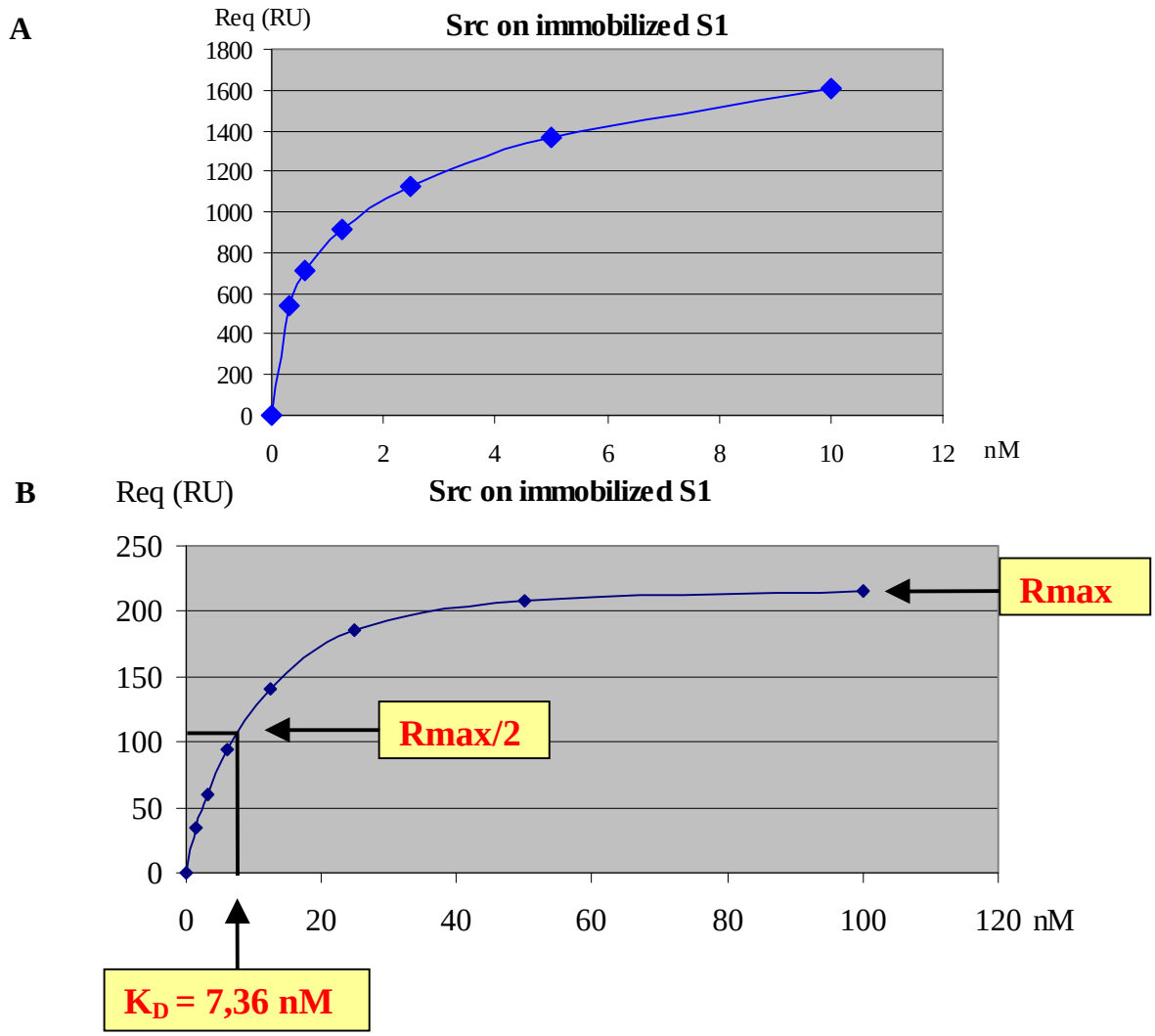
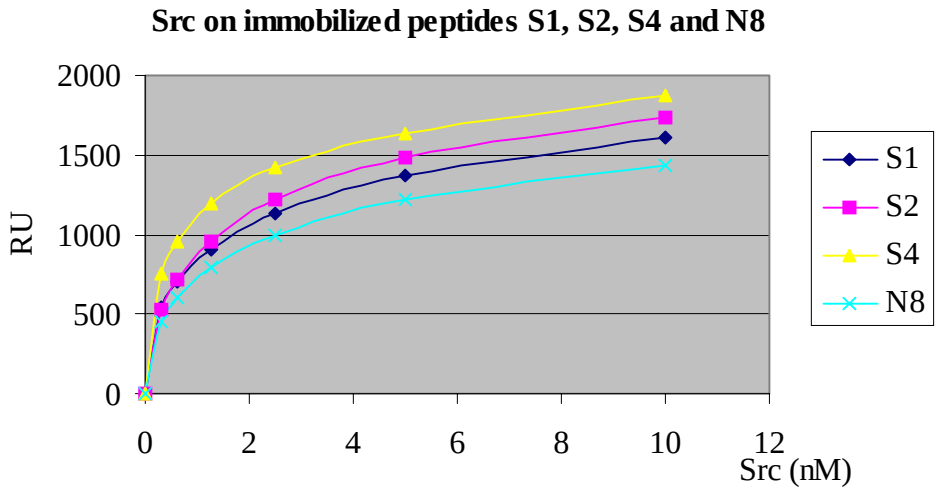


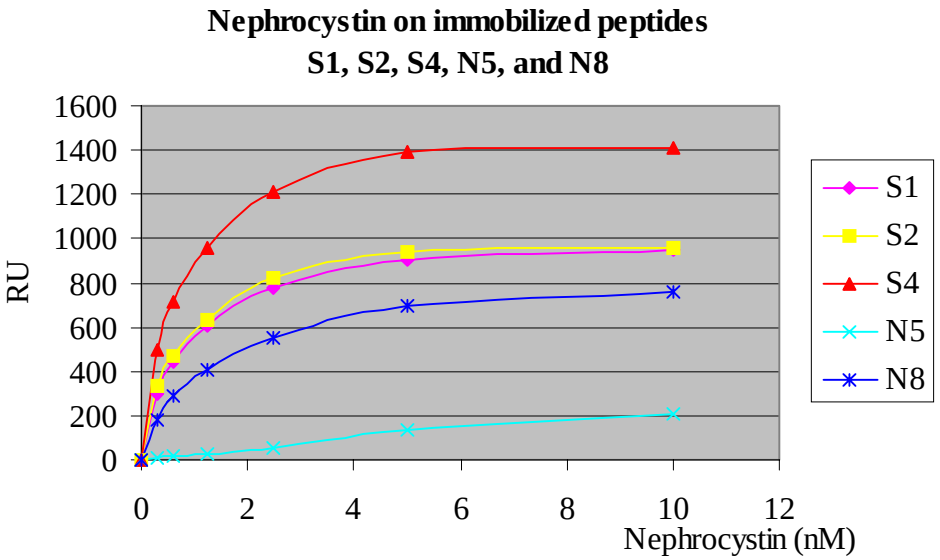
Figure III-50: Steady state affinity data evaluation of S1 clone. Plots of Req against Src concentrations, used to evaluate the K_D value of the S1 peptide/Src SH3 domains complex. (A-) Plot realized from the Req values obtained from Figure III-49, with Src concentrations from 0,3 to 10 nM. The equilibrium phase cannot be reached. (B-) Plot realized from another experiment, with Src concentrations from 1,5 to 100 nM. The Rmax value is deduced from the equilibrium phase. When Rmax/2 is reported on the curve, the K_D is directly readable, on the concentration axis.

As proved by our analyses, the S1 peptide binds to Src SH3 domain with a K_D of 7,3 nM. This is an exceptionally high K_D, concerning molecules involved in signal transduction. Indeed, SH3 modules must recognize ligands with high enough selectivity to maintain proper information flow, but low enough affinity, to allow for sensitive and dynamic modulation in response to changing signals. Typically, SH3 domains bind their targets with low affinities, as demonstrated by their usual K_D of 0,2 to 200 µM (Viguera et al., 1994; Cheadle et al., 1994; Rickles et al., 1994; Alexandropoulos et al., 1995). The recognition strategy of SH3 domains allows for high specificity binding, mainly through the prolines contained within the SH3 ligands, because there are no other natural sequences that can satisfy the minimal ligand backbone requirements (Nguyen et al., 1998).

A



B



C

	K_D (nM)	
	Src	Nephrocystin
S1	7,36	0.52
S2	nd	0,55
S4	nd	0,42
N5	nd	nd
N8	nd	1,05

Figure III-51: Steady state affinity data evaluation of Src and nephrocystin clones. Plots of Req against Src concentrations, used to evaluate the K_D value of SH3 domain/ligand complexes. (A-) Plot realized from the Req values of peptides S1, S2, S4 and N8, with Src concentrations from 0,3 to 10 nM. The equilibrium phase is not be reached. (B-) Plot realized from the Req values of peptides S1, S2, S4, N5 and N8, with Nephrocystin concentrations from 0,3 to 10 nM. The Rmax value is deduced from the equilibrium phase. When Rmax/2 is reported on the curve, the K_D is directly readable, on the concentration axis, as described in Figure III-50. (C-) K_D values obtained from the plots. The plot A cannot be used to evaluate the K_D, since the equilibrium is not reached. The K_D value of Src and S1, was calculated in Figure III-50/B. The K_D values for Nephrocystin SH3 domain, with S1, S2, S4 and N8, were calculated from the plot B. The K_D for N5 cannot be determined from this plot.

The same procedure was applied to the other SH3/ligand interactions. For each immobilized peptide, in contact with defined concentrations of Src or nephrocystin SH3 domains, the Req saturation responses were determined from the sensorgrams, and the plots Req versus SH3 concentrations were created. They are presented in Figure III-51/A and -51/B. The K_D were determined from these plots, as described above, and their values are shown in Figure III-51/C.

As previously explained, the K_D values for Src SH3 domain with S2, S4 and N8 cannot be determined from this experiment, since the Src concentrations used were not large enough to allow the obtention of an equilibrium state. However, the K_D between nephrocystin SH3 domain, and S1, S2, S4 and N8 have been precisely determined, and stand between 1,05 and 0,42 nM (Figure III-51/C), which is even more spectacular than the K_D of 7,36 nM determined between Src and the S1 peptide.

The K_D between nephrocystin and N5 peptide, cannot be determined from this experiment, the chosen concentrations did not allow the obtention of the equilibrium state.

H – Testing the functionality of the selected peptides, *in vivo*

We have characterized ligands against Src and Nephrocystin SH3 domains, using a new development of phage display techniques, the cosmix-plexing® technology. The isolated ligands interact *in vitro* with their target, with a very high affinity, as demonstrated by the phage ELISA, and the BIAcore real time interaction measurement analyses. The K_D of the various complexes, between Src or nephrocystin, and their peptides, are ranged from 7,3 to 0,4 nM, and are therefore, to our knowledge, the highest affinity ligands ever characterized for SH3 domains.

In order to have a better understanding of the signal transduction pathway in general, and of the functions of the SH3 domains in particularly, it is now crucial to evaluate whether these peptides are able to compete SH3 domain natural ligands, *in vivo*, and then, to study their influence on the action of SH3 domain containing proteins. The strategy chosen to deliver the functional peptides into the cells takes advantage of the HIV-1 Tat protein, which possesses the characteristic to cross cell membranes.

Src has essential functions in regulating cell growth and differentiation. However, little is known about Src role in signal transduction pathways and its downstream targets. Several knockout mice have been made to learn more about it. Src-knockouts in osteoclasts leads to

osteopetrosis. However, the absence of Src does not seem to disrupt any other cell activity. The subtlety of phenotype could in principle mean that Src family kinases play no important roles in the life of most cells in the body, but more likely, is due to redundancy, and compensation of function by other members of the Src family.

The most noticeable defect in Src^- mice is osteopetrosis (Soriano et al., 1991). Src is required for the normal differentiation and response of osteoclasts, a type of myeloid cell that absorbs bone. The Src^- osteoclasts have an autonomous defect and fail to form a "ruffling border" and resorb bone. The net result is a decreased absorption of bone, resulting in thickened bone, decreased marrow space, and also, in failure in tooth eruption. Osteopetrosis greatly increases mortality. Schwartzberg and coworkers (1997) developed a genetic system for studying Src function in the osteoclasts of transgenic mice by driving expression of wild-type and mutant versions of Src with the promoter of *tartrate resistant acid phosphatase (TRAP)*, a gene expressed highly in osteoclasts. They demonstrate that Src kinase activity is not essential for rescue of the *src*-/- phenotype. These results suggest that there are essential kinase-independent functions for Src in vivo, likely supervised by SH3 and/or SH2 domains.

In addition, Kaplan et al. (1995) studied the adhesion of Src^- fibroblasts to fibronectin. Src^- fibroblasts adhere and spread more slowly than control fibroblasts on a fibronectin surface, but adhere and spread normally on collagen. The adhesion and spreading defects were corrected by over-expression of either wild-type or kinase-defective Src. A kinase domain deletion of Src also promoted spreading, and required both SH2 and SH3 domains to do so. Requirement for SH3 domain may reflect the need for the molecule to localize to focal adhesions, but the requirement for the SH2 suggests an interaction with a phosphotyrosil protein.

In both osteoclasts and fibroblasts, Src appears to have a function which cannot be compensated by another SH3 domain, in the absence of Src (even if the fibroblasts defect, in the absence of Src does not seem to be dramatic). For both type of cells, it has been shown that the kinase domain of Src is not responsible of the defect. For the fibroblasts, it has been shown that both SH2 and SH3 domains of Src are required for normal adhesion and spreading of the cells.

Therefore, the creation of a cell model where Src SH3 domain natural ligands might be competed *in vivo*, by our isolated high affinity peptides, might be of great advantage to understand the particular role(s) of Src SH3 domain, in both osteoclasts, and fibroblasts. The Tat system appeared to be the best alternative to further investigate the function of Src *in vivo*.

It was decided to prepare Tat fusion proteins with each of the five clones which showed a high affinity to Src and/or nephrocystin SH3 domain on BIAcore measurements (S1, S2, S4, N5 and N8).

Dowdy developed an expression vector, to facilitate the cloning and purification of Tat fusion peptides (Nagahara et al., 1998). This pTAT-HA vector produces genetic in-frame TAT fusion proteins (Figure III-52). The vector pTAT-HA has an N-terminal 6 histidine leader, followed by the 11 amino acid Tat protein transduction domain, a hemaglutinin (HA) tag, and a polylinker. Steven F. Dowdy (University of Washington, Saint Louis, Missouri, USA) kindly provided us with his pTAT-HA vector, that we used for the cloning of our peptides.

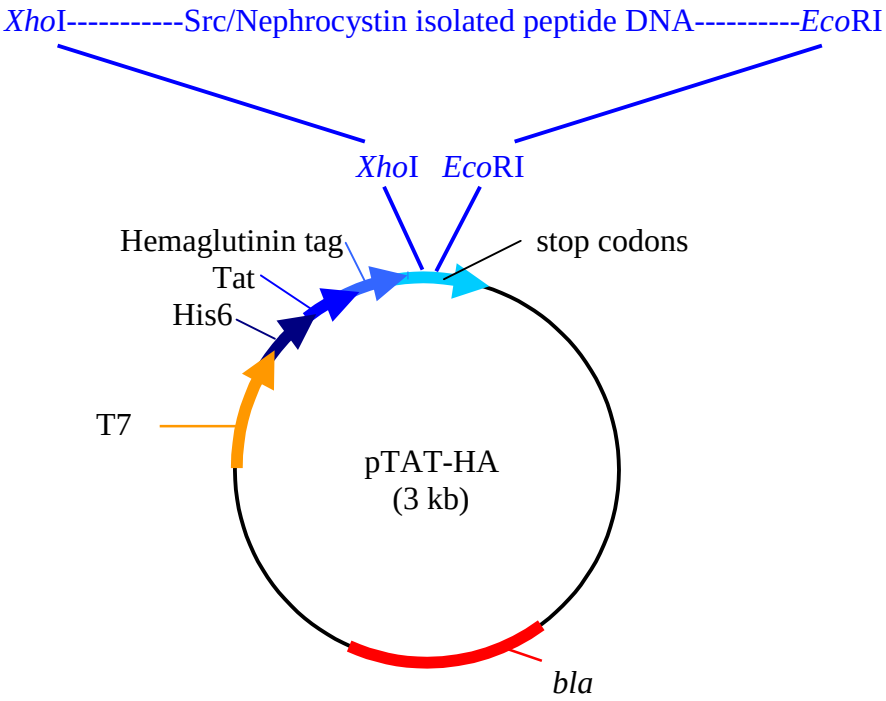


Figure III-52: Schematic map of the expression vector pTAT-HA. The 11 amino acid Tat protein transduction domain, is flanked N terminal by 6 histidines and C terminal, by the hemaglutinin tag, followed by the multiple cloning site. At the end of the cloning site, stop codons in all three reading frame are found. *Xho*I and *Eco*RI restriction sites were used to insert the Src and nephrocystin SH3 domain isolated peptide DNA. The Tat-fusion gene is under the control of the T7 promoter. *bla* stands for β-lactamase gene (ampicillin resistance).

The five peptides to be fused to the Tat proteins, have been directly amplified from pROCOS4/7, the plasmid vector which was used for the preparation of the CPLPPXP library. The primers used to this purpose were designed to introduce the *Xho*I and *Eco*RI restriction sites on both extremities of the peptides, to allow their insertion within the polylinker site of the pTAT-HA vector.

After the introduction of the oligonucleotides into the cloning site of the pTAT-HA, the vectors were transformed into BL21(DE3)LysS (Novagen), for the protein expression. Because of reduced structural constraints, high energetic, denatured proteins transduce more efficiently into cells than low energetic, correctly folded proteins. Once inside the cell, transduced denatured proteins are correctly refolded *in vivo* by host chaperones, like HSP90. Therefore, the proteins are expressed and purified in an urea-denaturing protein purification protocol (Nagahara et al., 1998).

500 ml culture of the BL21 cells containing the pTAT-peptides/cDNA is inoculated and incubated 4 hours at 37°C, pelleted, washed in PBS, and finally resuspended in a 8M urea buffer. After sonication, the expressed Tat fusion proteins are purified in Ni-NTA column (Qiagen), and eluted with 500 mM imidazole. Elution fractions of the purified pTAT-peptides are investigated on two SDS-PAGE (Figure III-53), one followed by staining with Coomassie blue, and the second analyzed in Western blot, to ensure that the observed bands correspond to the Tat fusion proteins (detection due to antibody anti-hemaglutinin).

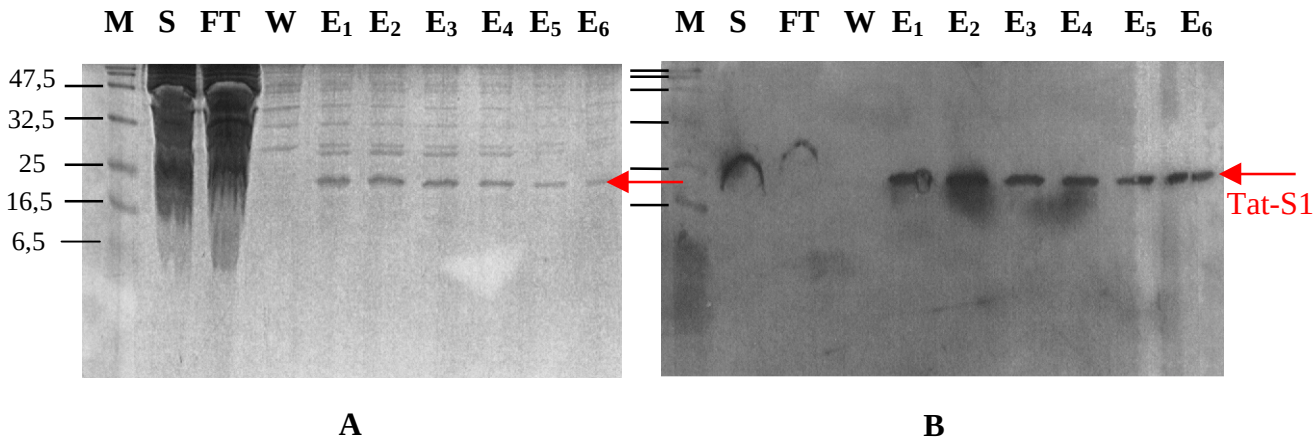


Figure III-53: Purification of Tat-S1 fusion protein, on a Ni-NTA column. The Tat-S1 protein loaded onto the column, was produced from 500 ml culture incubated 4 hours at 37°C, resuspended in a 8M urea buffer. After sonication, the expressed Tat fusion proteins are purified in Ni-NTA column. Washing of the column was performed with the previous 8M urea buffer, completed with 10 mM imidazole, and elution was achieved with 500 mM imidazole. (A-) Elution fractions of the purification of Tat-S1 analyzed by SDS-PAGE, followed by staining with Coomassie blue. S represents the supernatant of the cell extract, as applied to the column. FT is the flow through, W stands for a washing aliquot and E stands for the eluted fractions, which contain Tat-S1. The position and sizes (kDa) of molecular weight markers (M) are indicated. (B-) Western blot of a second SDS-PAGE, loaded with the same samples than the previous SDS-PAGE (A), run in parallel, and transfer to a nitrocellulose membrane. After blocking the membrane, the Tat-S1 fusion protein is detected with an antibody anti-hemaglutinin (Santa-Cruz). The red arrow, on both gels, indicates the position of the Tat-S1 fusion protein.

The western blot analysis shows that only a part of the Tat-fusion protein is retained by the Ni-NTA column, since a small amount of Tat-S1 is detected in the flow through. The loading cell extract contained a too high concentration of Tat-S1.

After purification, the urea of the buffer where the Tat fusion proteins were resuspended, had to be eliminated. One possibility to achieve this, is to dialyze the solution. However, this method leads to very poor yields. A good alternative, is an ionic exchange chromatography. After loading the sample and washing the SP column (Pharmacia), the Tat fusion protein was eluted with 1M NaCl, and finally desalted on a PD-10 disposable G-25 sephadex gravity column (Pharmacia) in PBS (Figure III-54). After the Ni-NTA column, the recovered Tat fusion proteins were not completely purified. However, the contaminants were not rescued by the SP column, and therefore, the final Tat fusion protein is a highly purified product.

To facilitate the detection of the Tat fusion proteins within the cells, the peptides were conjugated to the fluorochrom FITC. After determination of their concentration, with the Bradford technique, the proteins were ready to be transduced into cells. Unfortunately, time did not allow for analysis of the effect of these fusion peptides on Src, nephrocystin, or other potential cross reacting SH3 domains. However, an initial study of the capacity of our Tat-fusion peptides to reach the cytoplasm of the transfected cells was carried out.

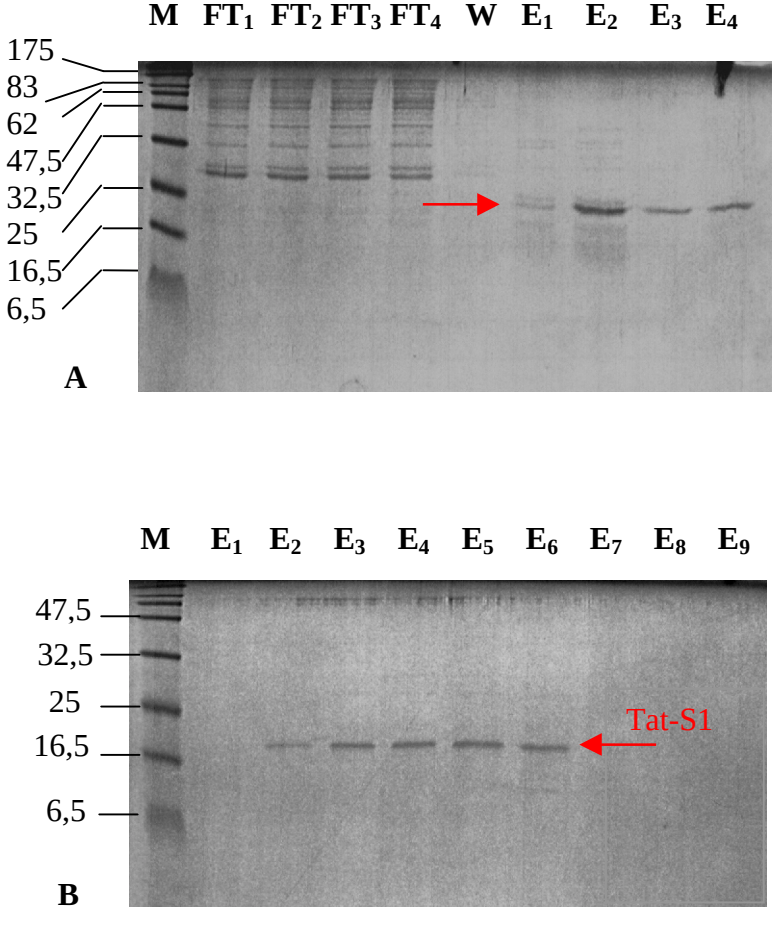


Figure III-54: (A-) Transfer of the Tat fusion proteins in an aqueous buffer, on an ionic exchange column. The Tat-S1 protein recovered from the Ni-NTA column is loaded in a SP column, in order to remove the urea. Flow through and elution fractions of the Tat-S1 are analyzed by SDS-PAGE, followed by staining with Coomassie blue. (B-) Recovered Tat-S1 fusion protein, after treatment on a desalting PD-10 column. The red arrow, on both gels, indicates the position of the Tat-S1 fusion protein. FT are the flow through, W stands for a washing aliquot and E stands for the eluted fractions, which contain Tat-S1. The position and sizes (kDa) of molecular weight markers (M) are indicated.

Prior to further analyses, the ability of the Tat fusion peptides to interact *in vitro* with the Src and nephrocystin SH3 domains was tested. Therefore, ELISA assays were performed to determine whether the FITC-conjugated-peptides were still able to bind to SH3 domains. Various concentrations of Tat fusion peptides were incubated with immobilized Src and nephrocystin SH3 domains, after blocking of the microtiter wells. The amount of bound FITC-Tat peptides was evaluated by estimation of the fluorescence detected in the wells after intensive washing steps, on ELISA reader, sensitive to fluorochroms. This demonstrated the ability of the Tat fusion peptides to bind specifically to the Src and nephrocystin SH3 domains.

NIH-3T3 cells were chosen for the *in vivo* study of Tat fusion peptide delivery, due to the fact that Src plays an important role in fibroblasts, and to their ease of manipulation. To analyze the ability of Tat fusion proteins to transduce into cells, purified Tat-S4 fusion peptide was conjugated to fluorescein (FITC), and added to the culture media of NIH-3T3 fibroblasts (160 nM of Tat-S4 directly added to 10^6 NIH-3T3 cells). After one hour incubation, the cells were trypsinized, collected, resuspended in fresh media/FCS (to reduce the background), and analyzed by fluorescence activated cell sorting (FACS). However, no fluorescence was detected.

Currently, the most likely hypothesis for the Tat translocation mechanism, involves a tight ionic interaction between the basic groups of the Tat side chain, and the negative charges of the phospholipid heads, which would induce a local invagination of the plasma membrane. The local reorganization of the phospholipid bilayer would then lead to the formation of inverted micelles with the peptide enclosed in the hydrophilic cavity and ultimately to the cytoplasmic release of the peptide. This might mean that flexibility of the cell membrane is required. In addition, Dowdy warned us that nonadherent cells are best experimentally.

Therefore, a hypothesis was conceived, that NIH-3T3 adherent cells were not appropriate to a Tat-fusion peptide transduction test, because the adhesion of the cells to the plastic material might stretched the membrane to such an extent, that the plasma membrane might not be able any longer to form a local invagination, and therefore, to trap the Tat-fusion peptide.

Thus, another method was developed to deliver the Tat-fusion peptide within the NIH-3T3 cells. NIH-3T3 cells were trypsinized, and laid on an untreated plastic material, where

they have no possibility to adhere. The FITC-Tat fusion peptides were added to this suspension of NIH-3T3 cells. At various incubation times, some cells were resuspended in fresh medium, and analyzed by FACS. By using this strategy, the transfection of Tat-fusion peptides into the cells was successful.

After one hour of incubation between the cells and the Tat fusion peptide, cell aliquots were transferred into fresh culture flasks, where the cells were finally allowed to adhere to the surface. These flasks were assigned for confocal microscopy analysis.

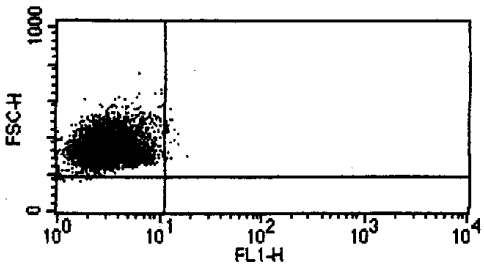
It is crucial to let the cells adhere, after a short incubation time with the Tat-peptides where adhesion is prevented, otherwise, they aggregate, and die. However, no alteration of the cell morphology, or growing characteristic has been observed when the cells were prevented to adhere for a short time.

The results of the FACS analysis are presented in Figure III-55. The kinetic analysis of FITC labeled Tat-S4 peptide (160 nM) directly added to NIH-3T3 cells was performed at 0, 15, 30, 45 and 60 min. As shown in Figure III-55/A and -55/B, after one hour, nearly 80% of the cells are efficiently transduced with the Tat fusion peptide.

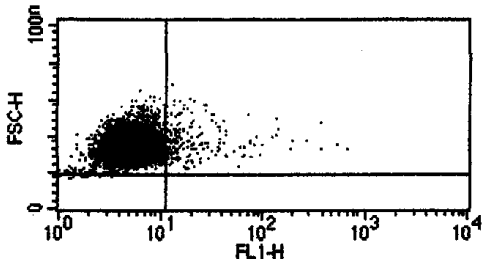
A dose dependent assay of cells treated for one hour with 20, 40, 80 and 160 nM of Tat-S4 fusion peptide (Figure III-55/C) did not allow the demonstration of a concentration dependency for protein transduction, in disagreement with Dowdy observations (Nagahara et al., 1998). It might have been interesting to repeat the FACS analysis with a longer incubation time (2 hours), and higher concentrations of Tat fusion peptide. The results are showing that it is possible to efficiently transduce a very high proportion of the NIH-3T3 cells with Tat-fusion peptides.

Figure III-55: Analysis of transduced proteins. 10^6 NIH-3T3 cells were used for the transfection. They were placed on an untreated plastic material, where they have no possibility to adhere. Before FACS analysis, the aliquot of cells to be analyzed, was pelleted and resuspended in fresh medium to eliminate the background. (A and B-) FACS kinetic analysis of 160 nM FITC labeled Tat-S4 fusion peptide added to NIH-3T3 cells, at 0, 15, 30, 45 and 60 min. (A-) FACS profile of the cells were FITC was detected. (B-) Percentage of cells containing FITC labeled Tat-S4 fusion peptide after 0, 15, 30, 45 and 60 min. (C-) FACS dose analysis. Percentage of cells containing FITC labeled Tat-S4 fusion peptide after one hour incubation, when 0, 20, 40, 80 and 160 nM of Tat fusion protein are added to the medium of NIH-3T3 cells.

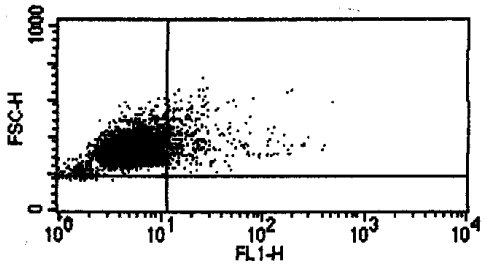
A



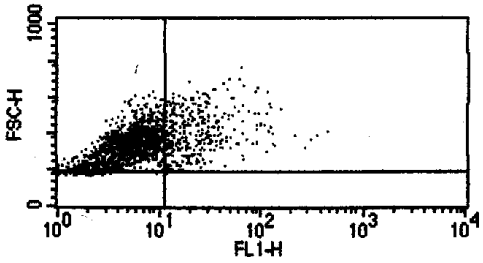
NIH-3T3 control
Without FITC-Tat-S4
Background: 0,5%



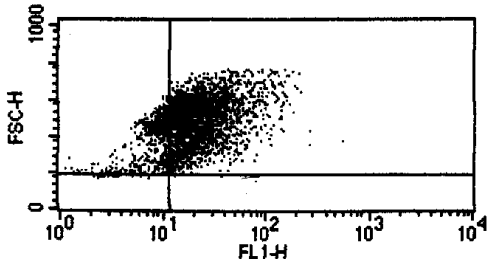
After **15 min**
6,5% of cells are transduced



After **30 min**
10,3% of cells are transduced



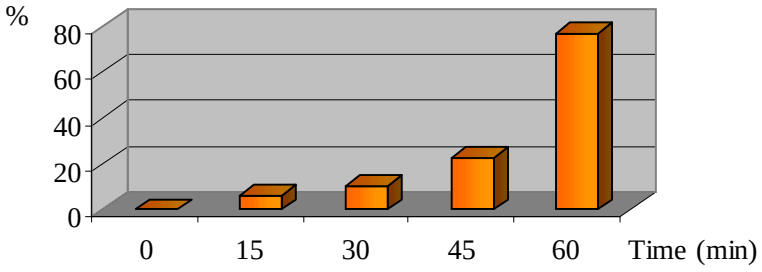
After **45 min**
22,8% of cells are transduced



After **60 min**
77% of cells are transduced

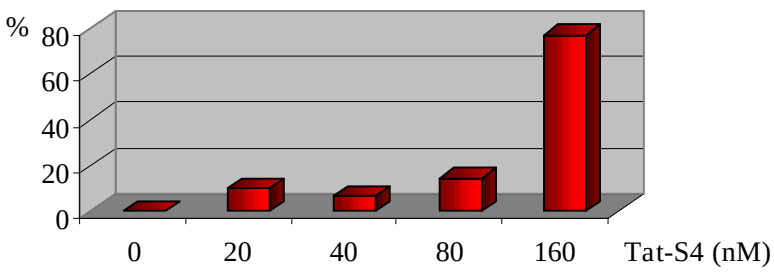
FACSAN kinetics:
percentage of incorporated Tat-S4 into NIH-3T3

B



FACSCAN: percentage of incorporated Tat-S4
into NIH-3T3, according to Tat-S4 concentration

C



Confocal microscopic analysis showed the presence of Tat-fusion peptides within the cytoplasm, but not in the nucleus (Figure III-56/A). However, in the work performed by Dowdy and coworkers (Nagahara et al., 1998; Vocero-Akbani et al., 1999), the Tat-fusion proteins were found in both cytoplasmic and nuclear localization. It has been shown that the 11 amino acid Tat translocation domain used for the protein fusion, contains a nuclear localization signal sequence. The absence of Tat-S4 within the nucleus means that it is retained in the cytoplasm, and a hypothesis would be that Tat-fusion peptides are maintained within the cytoplasm through their interactions with SH3 domains.

Tat-fusion peptides transduced both healthy and dead cells, as seen in Figure III-56. But curiously, the incorporation profiles look very different. In healthy cells, the Tat-S4 fusion protein is found ubiquitously in the cytoplasm (Figure III-56/A). However, in dead cells, it remains attached on the border of the cell, close to the cell membrane (Figure III-56/B).

Src family kinases are found associated with cell membranes, both the plasma membrane and intracellular membranes, such as the endoplasmic reticulum and endosomes. The Src N-terminal 7 residues are necessary and sufficient to program Src for myristoylation, which is necessary for the membrane localization. In addition to the myristoylation signal, Src has basic residues in its membrane localization region (SH4), which contribute to membrane association, most likely by interaction with the acidic headgroups of membrane phospholipids.

It seems that in dead cells, the Src are directed to the cell membrane. Does it reflect an absence of activity of the cell, i.e. Src is not necessary anymore at the endoplasmic reticulum or other intracellular membranes? Or does the cell send a signal to the Src protein (and related proteins) to localized at the cell membrane?

The NIH-3T3 cells observed on confocal microscopy were kept, and checked once more every day, to evaluate how long it was possible to detect the FITC-Tat fusion protein within the cells. It has to be emphasized that the transfection is not permanent, the Tat-peptides are able to go in and out the cell, through the membrane. Therefore, the observation that after several days, and regular changes of medium/FCS, FITC labeled Tat-peptides were still detected within the NIH-3T3 cells, reinforces the previous hypothesis that Tat-fusion peptides are maintained in the cell cytoplasm, likely through their interactions with SH3 domains. This hypothesis is further supported by the observation that it was impossible to detect any trace of FIT-Tat-peptide within the nucleus of transfected cells.

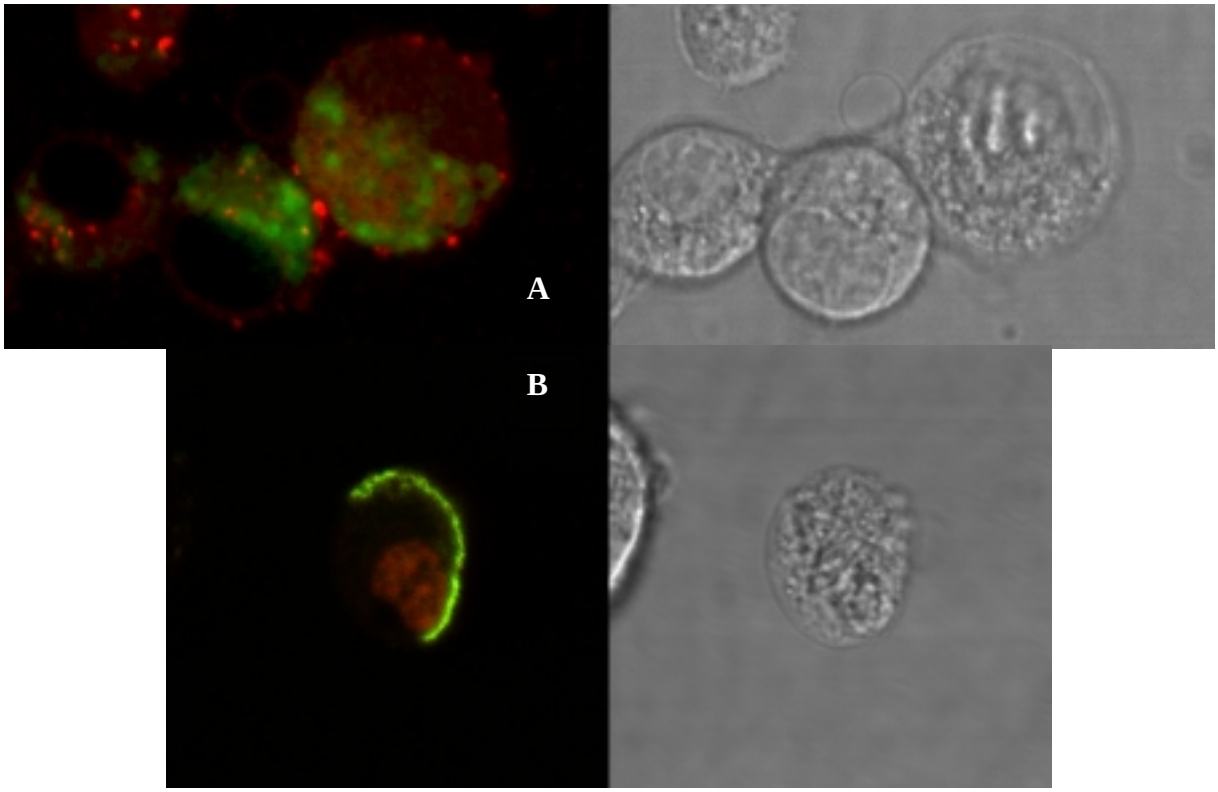


Figure III-56: Fluorescence confocal microscopy sections of NIH-3T3 treated with 160 nM of FITC labeled Tat-S4 for one hour. Before the observation under confocal microscope, the cells were washed with fresh medium/FCS, to remove the background. Propidium iodide has been added to the medium, to check for cells mortality, since dead cells incorporate propidium iodide in their nucleus. Fluo spheres (0,04 µm) provided by K. Dittmar were used as positive control, since they are able to transfect cells. Both propidium iodide and fluo sphere emits fluorescence in the red light. (A-) Healthy NIH-3T3 cells which incorporated both FITC-Tat-S4 (green) and fluo sphere (red) in their cytoplasm. (B-) Dead NIH-3T3 cell. Propidium iodide is incorporated in the nucleus, and FITC-Tat-S4 fusion peptide is maintained at the membrane level of the cell.

In summary, we succeeded in isolating and characterizing very high affinity ligands for Src and Nephrocystin SH3 domains. When fused to the 11 amino acid Tat transduction domain, the peptides are able to transfect the cells, even when FITC labeled, and seem to be able to interact *in vivo* with local SH3 domains. Collaborating laboratories might carry on the work with these Tat-fusion peptides, to investigate their role(s) *in vivo*, on SH3 domain containing proteins. It might be possible to elaborate cellular model where signal transduction pathways could be tightly controlled and regulated. These models would find numerous applications in medical and pharmaceutical domains.

IV - DISCUSSION

This thesis aimed at establishing a method leading to the isolation and characterization of very high affinity ligands for SH3 and EVH1 domains. Two strategies based on biased phage libraries were experimented to reach this aim: the use of constrained peptide libraries, and of cosmix-plexing® peptide libraries. The first did not allow identification of any SH3 ligands. However, the cosmix-plexing® library permitted the isolation of SH3 domain ligands presenting a K_D approximately 100 to 1000 fold lower than previously characterized SH3 ligands.

A- Constrained libraries

Four biased constrained libraries were prepared in the phagemid vector pSKAN8, where the hypervariable domains replaced a 7 amino acid loop (residues 17 to 23) of PSTI (Pancreatic Secretory Trypsin Inhibitor). The constrained peptides were essentially monovalently displayed on the surface of the phage particles, as a pIII-fusion protein.

Chen et al. (1993) first reported the use of a biased library which contained three fixed prolines and was conceived to select ligands for the PI3K SH3 domain. This study demonstrated that biased libraries yield peptides with affinities that were among the highest reported for SH3 ligands (K_D of 7 to 30 µM), and comparable to those of peptides derived from natural ligands.

The four constrained banks pSKAN8-HypD, -HypE, -HypF and -HypG displayed peptides bracketed by two cysteines, and therefore, presented as a loop on phage surface. The periplasm of *E.coli*, where the phage particles are assembled, is an oxidizing environment and favors the formation of disulfide-bridged cysteines. The presence of disulfide-mediated peptide loops on phage has been demonstrated experimentally (Hart et al., 1999). The displayed peptides were as followed:

HypD	*Cys*	$(Xxx)_4$	**Pro**	Xxx	**Arg**	*Cys*		
HydE	*Cys*	**Arg**	$(Xxx)_3$	**Pro**	$(Xxx)_2$	*Cys*		
HypF	*Cys*	$(Xxx)_2$	**Pro**	Xxx	Xxx	**Pro**	$(Xxx)_2$	*Cys*
HypG	*Cys*	$(Xxx)_3$	**Pro**	**Pro**	Xxx	**Pro**	$(Xxx)_2$	*Cys*

The fixed prolines or arginines gave a selection advantage for SH3 domains which show preferences for proline rich ligands, containing an arginine before, or after the proline core motif.

The diversities obtained for the libraries, range between $1{,}8 \times 10^8$ and 9×10^9 transformants

(Table III-2), sizes which are sufficient to represent most of possible random peptide combinations.

Most small peptides that bind specifically to target proteins *in vivo* are constrained, commonly by disulfide bridging. Phage displayed libraries have been used to isolate high affinity cyclic ligands for several proteins, including integrins (Pasqualini et al., 1995), the erythropoietin receptor (Wrighton et al., 1996), the Grb2 SH2 domain (Oligino et al. 1997), calmodulin (Pierce et al., 1996), and syntrophin PDZ domains (Gee et al., 1998). In most cases, these peptide ligands do not resemble the primary sequence of the natural interacting proteins. Instead, they act as structural mimics (Gee et al., 1998).

The use of libraries of constrained peptides have both advantages and disadvantages. Constraining peptides might increase their affinity, and perhaps specificity, but constraining peptides into an incorrect conformation would probably prevent binding (Clakson and Wells, 1994). Imposing constraints on the displayed peptides can confer a number of important advantages, a few of them are described here.

A constrained peptide can show higher affinity for a target than an unconstrained peptide (Ladner, 1995). A ligand that has the shape to complement a binding site should have higher affinity than a similar ligand that has high entropy when it is separated from the target. The more tightly a peptide segment is constrained (for some small stable constrained peptides, the structure may be so rigid that only side groups have significant freedom; Cunningham et al., 1994) , the less likely it is to bind to any particular target. However, if it does bind, binding is likely to be tighter and more specific.

In addition, a constrained peptide can present exposed residues that may be in an unfavorable conformation within a linear peptide. For example, hydrophobic parts of side groups tend to become buried unless they are constrained to remain on the surface.

An important feature of constrained peptides, is that their binding properties are transferable between phage-bound protein and phage-free protein to a much greater extent than those of linear peptides (Dennis and Lazarus, 1994; Roberts et al., 1992). Therefore, further work with synthetic peptides might be more successful, since more faithful, in the case of constrained peptides.

Constrained peptides are more stable than linear ones, under many chemical and biological conditions.

Peptides containing disulfide-closed loop, selected for binding to sera of patients (Folgori et al., 1994) or to monoclonal antibodies (Meola et al., 1995) have been found to be

better immunogens than are linear peptides.

Taken together, these data show that isolating ligands from constrained libraries provides good chances to identify high affinity and specificity ligands, if the constraints imposed on the displayed peptides, are advantageous for the interaction with the target. However, using constrained library to isolate ligands against SH3 domains was not a project without risks, since at the beginning of this work, it was acquired that SH3 ligands prefer or require a polyproline type II helix, to be able to interact with the SH3 domains.

In the light on these data, the unsuccessful panning selections on SH3 domains were not entirely surprising.

It has to be emphasized that other groups have reported difficulties in screening for SH3 binding molecules, using constrained libraries, however, not biased (Chen et al., 1993; Cheadle et al., 1994; Yu et al., 1994; Kay and Quilliams, personal communication). To our knowledge, only one group succeeded to isolate a cyclic Src SH3 ligand, by using a “mirror image phage display” technology which allows characterization of D-peptide ligands (Schumacher et al., 1996).

By using this “mirror image” phage display method, Schumacher et al. (1996) succeeded to isolate non proline rich ligands against Src SH3 domain. In this technique, a D-enantiomer of the Src SH3 domain was synthesized, and used to isolate L-ligands, from a biased phage library ($S/CX_{10}S/C$). The synthesized D-enantiomeric forms of the L-peptide ligands were expected to interact with the natural L-Src SH3 domain (Figure IV-1). This method was developed to overcome problems linked with naturally occurring L-peptides within cells: D-peptides provide useful starting points for the design or selection of novel drugs, since opposite to the L-peptides, they are not subject to degradation by naturally occurring enzymes. Due to their sensitivity to cellular proteases, peptides composed of L-amino acids can be efficiently digested by antigen presenting cells, and exposed on their surface, with the major histocompatibility complex class II. As a result, they are presented to the T helper cells which induce an immune response that impairs the activity of L-peptide drugs.

The two L-ligands isolated with this technique against the D-SH3 domain do not resemble the usual SH3 ligands, and are bracketed by cysteines. But since it is a D-enantiomer, the sequences are not comparable. An unique cyclic D-enantiomer peptide (CLSGLRLGLVPC) was shown to bind *in vitro* to the L-Src SH3 domain with a weak dissociation constant of 63 µM, however, only when the disulfide bridge was maintained.

Since the disulfide bridge imposes constraints on the peptide, its structure is not comparable with usual SH3 ligands.

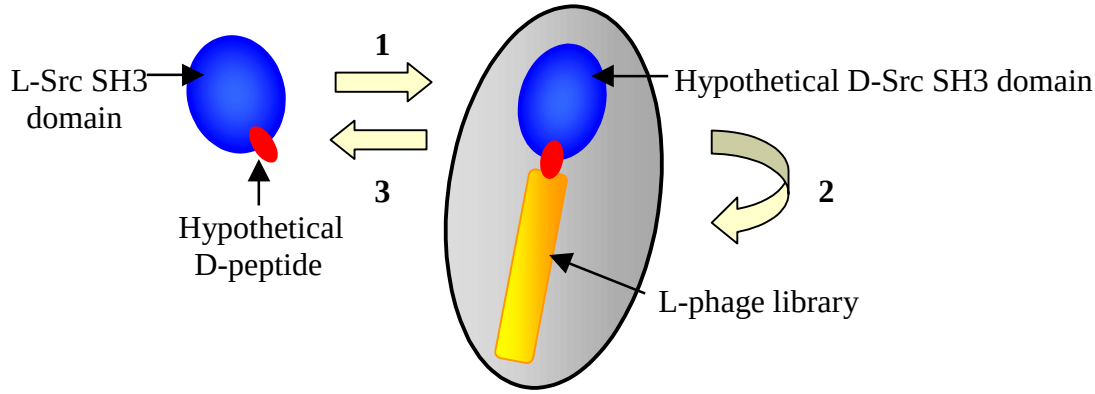


Figure IV-1: Illustration of the mirror image phage display technique, according to Schumacher et al. (1996). A D-enantiomer of the Src SH3 domain is synthesized (1), and used as target for screening with a phage library (2). A D-enantiomer of the isolated peptide interacting with D-Src SH3 is synthesized (3), and interaction with the L-Src SH3 domain is tested.

Greater success was expected with the constrained pSKAN8 libraries, since they were biased, and that the introduction of fixed proline(s), or structural motifs within the sequence, should have increased the average affinity of the ligands for the SH3 domains. In addition, our libraries displayed longer peptides on the surface of the phage (7 to 9 amino acids, between the two cysteines, whereas Chen et al. (1993) used a 5 amino acid long peptide library, and Cheadle et al. (1994), 6). This should have increased our chances of more completely covering the SH3 binding site, and in addition, to give a larger freedom to the displayed peptides, since the larger the loop, the less the constraints.

Another reason for pessimism with this approach, was the risk that the constraint imposed on the peptides due to the disulfide bridge which takes place between the two cysteines (on both extremities of the hypervariable domain), prevented the formation of the polyproline type II (PPII) helix, essential for ligand binding to SH3 domains. The residues of a disulfide-closed loop are only partially constrained by the disulfide bridge, and the amino acid sequence within the loop can impose additional constraints (Cunningham et al., 1994). However, it is not known whether these additional constraints can be as complex as a helix and, in our case, as a PPII helix.

Furthermore, it is admitted that disulfide-closed loop sequences that are rich in prolines, valines, or isoleucines are more constrained than others (SH3 ligands are prolines rich, but also commonly possess Ile between its prolines), further reducing the chances for our

displayed peptides, to form a PPII helix (Ladner, 1995).

The experiments with SH3 ligands bracketed by cysteines, and expressed on the surface of the phage particles, at sites where the hypervariable domains were located, and the four constrained libraries generated, tend to demonstrate the impossibility for peptides constrained by disulfide bridges, to adopt a PPII helix conformation (section III-B-4).

The SH3 ligand used for this experiment was chosen on the basis of work carried out by other groups. Brian Kay's group isolates ligands for various SH3 domains, and investigates them for cross-reaction (Sparks et al., 1996). The selected SH3 ligand was originally identified against the N terminal SH3 domain of Grb2 (Grb2N), but showed strong cross-reaction on phage ELISA, with other SH3 domains, including Src, Yes, and the N terminal SH3 domain of Nck (NckN).

Two forms of this SH3 ligand were prepared, which only varied by the size of the displayed peptides (9 and 18 amino acids between the cysteines, respectively). Both forms were used to check if the number of residues between both cysteines, influences or not, the formation of the polyproline type II helix, since it is known that a longer constrained peptide has more opportunities to form additional constraint. The sequences of the short and long SH3 ligands used in the test, were as followed:

Cys Pro Tyr **Tyr Gln Arg Pro Leu Pro Pro Leu Pro** Leu Ser His Phe Glu Ser Val *Cys*

Short SH3 ligand (bracketed by cysteines)

Long SH3 ligand

On phage ELISA, neither the short nor the long SH3 ligands were able to interact with any of the various tested SH3 domains (Src, Grb2, Grb2N and NckN). This suggests that the presence of both cysteines, and the resulting disulfide bridge (*E.coli* periplasm, where the phage particles are assembled, is an oxidizing environment), obstructs the interaction between the SH3 domains and the tested SH3 ligands, likely by preventing the formation of the required PPII helix.

However, a positive control is missing to clearly demonstrate the hypothesis that disulfide bridges within short peptides, obstruct the formation of polyproline type II helices. It would have been necessary to check if the original SH3 ligand, that was tested in this experiment, was also able to bind, in our hands, to the Src, Grb2N and NckN SH3 domains, in the absence of cysteines. Another positive control could have been to test if the presence of the reducing agent DTT allows interaction between short or long SH3 ligands and SH3

domains (Gee et al., 1998).

One could think that another cause of disturbance of the formation of the PPII helix, would be if the immediate environment on the surface of the PSTI, interferes with the hypervariable domain. However, this should not be the case. The hypervariable region replaced a 7 amino acid loop within the PSTI. In the natural PSTI, two cysteines bracketed this 7 amino acid fragment and impose a constraint on this region, resulting in a loop. PSTI is 56 amino acid long, of which 6 amino acids are cysteines. Except for the random peptides which would contain an extra cysteine, the equilibrium of the protein is conserved, and the cysteines flanking the variable domain are unlikely to form disulfide bridges with one of the four other cysteines present in PSTI. Furthermore, PSTI also does not interact with the pIII phage protein to which it is fused, as shown by Röttgen and Collins (1995). In their work, the pSKAN8 vector was successfully used to display a 7-8 residue peptide within the PSTI loop, on the surface of a phage. The resulting library permitted the isolation of ligands for chymotrypsin and elastase.

B- Cosmix-plexing® library: CPLPPXP

The second approach, based on developing a cosmix-plexing® library, allowed the identification of ligands for Src and nephrocystin SH3 domains, but also for EVH1 domains. The Src and nephrocystin SH3 ligands show excellent K_D value 100 to 1000 fold lower than previously reported SH3 ligands (in the range of nM, instead of µM). The advantage of cosmix-plexing® libraries, is that recombination between left and right parts of the hypervariable domain is induced by simple digestion with type IIs restriction enzyme, which recognizes a defined DNA sequence, in the constant region flanking the hypervariable domain, but cuts a fixed distance away from the binding site, within the hypervariable region. Religation of both parts of the variable domain, leads to the creation of new clones, allowing dramatic increase of the library diversity, and optimization of the whole region.

The cosmix-plexing® library generated for the screening on SH3 and EVH1 domains, the CPLPPXP library, is a biased peptide library, displaying essentially monovalently, a 13 amino acid long peptide as a linear prolongation of the pIII protein, on the surface of phage particles. The library was constructed in two steps (Figure III-11), in the phagemid vector pROCOS4/7. First, a synthetic hypervariable oligonucleotide encoding for a 8 amino acid random peptide, was cloned into pROCOS4/7. Then, the second stage was the introduction of a project

specific cassette, encoding for a proline rich motif, within the previous 8 amino acid peptide insert.

The primary libraries have been prepared in four sets, differing only by the dinucleotide corresponding to the cohesive ends, generated upon cleavage by the type IIs restriction enzyme (at the ZZ position, in Figures III-11, and III-12). This ZZ position was designed to prevent head to head joining of the cohesive ends, during religation, to maintain the right orientation of the fragments. In addition, when the four libraries were pooled, the NZZ codon allowed presentation of all amino acids, except cysteine.

The four primary banks together, reached a total diversity of $2x10^8$ transformants. Since the libraries were recombined together by cosmix-plexing®, to introduce the proline rich cassette within the 8 amino acid insert, many new clones were created, and the diversity has been increased. The quality of the four banks have been evaluated, and was shown to be sufficient to continue with insertion of the project specific cassette. For each of the libraries, around 80% of the clones were properly formed (Table III-9), and these clones present an amino acid frequency correlating with the expected values (Table III-10). The four primary libraries were pooled before introduction of the proline rich cassette.

The project specific cassette was designed to introduce a proline rich core motif within the hypervariable domain, so as to improve the possibilities of finding SH3 and EVH1 domain ligands.

16 sets of cassettes were prepared, which differed only at the ZZ dinucleotide positions, at both extremities of the cassettes (Figure III-11/B), i.e. the sites used for insertion. During insertion of the cassette within the 8 amino acid hypervariable stretch where the ZZ dinucleotides varied at its right and left ends, formation of concatemers were induced by ligation at high DNA concentration. The concatemers were required to recombine both parts of the hypervariable domains, and to increase the diversity of the bank. The concatemers were resolved into single phagemid molecules on cleavage with a restriction enzyme recognizing an unique region within the vector, followed by ligation at low DNA concentration, to favor formation of monomers.

The diversity obtained for the CPLPPXP library was equivalent to $2{,}3x10^8$ transformants. The principle of cosmix-plexing®, however relies on the fact that the diversity is increased each time a cosmix-plexing® is performed. The size of the library was therefore thought to be sufficient. In addition, the frequency of the amino acids was very similar to the expected values of a NNB library (Table III-12).

Analysis of randomly picked clones revealed that 91% of the clones were correctly formed (having continuous open-reading frame), and thus, adequate for panning selections. These correct clones formed two groups. 71% of the library displayed the proline rich 13 amino acid long peptides on the surface of the phage (Table III-11). The other 20% of the clones, were religated without the cassette, and therefore, only the 8 amino acid entirely random peptide of these clones, was exposed on the surface of the phage. The formation of these latter could have been avoided, by dephosphorylation of the ends of the cleaved vector before ligation with the cassette. However, this would have resulted in a reduction of the efficiency of the subsequent transformations, i.e. an unacceptable reduction in library diversity. The two types of displayed peptides were as followed:

Xxx_5 **Pro Pro** Xxx **Pro** Xxx_4	71% of CPLPPXP
+	
Xxx_8	20% of CPLPXP

The CPLPPXP library is convenient to carry on panning selection. Due to the fixed prolines, this cosmix-plexing® library has a strong selective advantage to isolate ligands against the SH3 and EVH1 domain targets.

1- SH3 domains

Several groups previously successfully isolated ligands against various SH3 domains. Originally, two proteins that bind specifically to the Abl SH3 domain, were isolated by screening a λgt11 cDNA library with a GST-Abl fusion protein (Cicchetti et al., 1992). One of these proteins, 3BP-1, contains a region of similarity to GAP-rho, the guanosine triphosphate-activating protein for the Ras-related protein rho. Because of this similarity, it has been thought that 3BP-1 could function as an intermediary between SH3 domains and the signal transduction pathway mediated by the Ras-related guanosine triphosphate (GTP). Further work on these two proteins (3BP-1 and 3BP-2) allowed the localization of the SH3 binding sites to a 10 amino acid stretch very rich in proline residues (APTMPPPLPP, for 3BP-1; Ren et al., 1993). Individual positions were mutated to alanine (alanine scanning), to determine which residues were important. The positions P2, P7, and P10 (referring to the positions of the amino acid in 3BP-1), but also A1, and P9, seem crucial to binding affinity with Abl SH3 domain (APxxxxPxPP). A GenBank data search led to the identification of some other peptides showing affinity for Abl SH3 domain, and alignment of their binding sites showed a clear consensus sequence: xPxxPPPΨxP (where Ψ represents hydrophobic residues; Ren et al., 1993).

Taken together, these data showed the importance of the proline core motif for peptides to interact with Abl SH3 domain, but also with Src SH3 domain, since some cross-reactions were noticed. Random peptide libraries (Sparks et al., 1994; Rickles et al., 1994) and screening of expression libraries (Viguera et al., 1994; Alexandropoulos et al., 1995) were then extensively used against various SH3 domains, and confirmed the crucial role of proline rich core motif for interaction between SH3 domains and their ligands.

Introduction of the use of biased libraries to characterize SH3 ligands (Rickles et al., 1994; Yu et al., 1994; Rickles et al., 1995; Feng et al., 1995; Sparks et al., 1996), confirmed the essential need of a proline core motif, but furthermore enabled the determination of more information about structural requirements of the proline motif flanking regions, within SH3 ligands. The amino acid preferences of the flanking regions allowed the classification of isolated ligands into two classes, I and II, which bind the SH3 domain in opposite orientations (Feng, et al., 1994; Yu, et al., 1994; Lim et al., 1994). The class I ligands show the consensus sequence **R**XLPPXP, while the class II ligands present as consensus sequence XPPLPX**R**. The major difference between both classes is the position of a critical arginine (in bold), which forms a salt bridge with a conserved acidic residue in the SH3 domain, and thus, determine the orientation of the ligand. Class I and II ligands bind in NH2 - COOH and in COOH - NH2 orientation respectively.

SH3 domains have been shown to bind to ligands either of class I or II (Sparks et al., 1996). Only Src, PI3K and Grb2N appear to be able to bind to ligands belonging to both classes (Feng et al., 1994; Yu et al., 1994). The Table IV-1 classified the SH3 ligands, according to their ligand class preferences.

Class I ligands	Class II ligands
Src **PI3K** **Grb2N** Yes Abl	**Src** **PI3K** **Grb2N** Cortactin p53bp2 PLCγ CrkN

Table IV-1: Classification of SH3 domains, according to the type of ligands that they bind. The ligands were assigned to be class I or II, with respect to the position of an arginine, before (I), or after (II) the proline core motif (Sparks et al., 1996).

SH3 ligands are centered on a proline core motif, which adopts the PPII helix conformation, and interacts with the aromatic pockets of SH3 domains. Peptides with PPII conformation form an extended molecular scaffold that is well suited for the shallow groove of the SH3 ligand-binding site (Yu et al., 1994). PPII helices on globular proteins (natural SH3 ligands) are ideal mediators of protein-protein interactions since most of these motifs are located on the surface of their proteins (Adzhubei and Sternberg, 1993).

The proline core motif is flanked by N and C terminal amino acids, one of each being assumed to be responsible for the specificity of each SH3 domain. The binding of the flank appears to require the presence of the core, since a peptide having only the flank sequence does not bind SH3 domain (Feng, et al., 1995). The core is assumed to serve as a common anchor used by SH3-binding peptides for initial docking, while the flank allows more extensive contact at the intermolecular interface, thereby providing increased affinity and specificity.

The ligand binding site of SH3 domain, consists of three pockets that together form a relatively shallow groove on one side of the molecule (Feng et al., 1994). Two hydrophobic pockets (n-Src and RT-loop) accommodate the proline core motif of SH3 ligands, whereas the specific pocket (formed by side chain of Asp^{99} and Trp^{118} in Src SH3 domain), accommodates the SH3 ligand arginine residue (Feng et al., 1994).

1-1 Src SH3 domain

The idea to use Src SH3 domain as model to optimize the affinity selection protocol was based on results previously achieved in other laboratories. Src is probably the SH3 domain which was, until now, the most studied, and the Src ligand preferences were already quite precise before the beginning of this thesis. This was considered as an advantage, since it was expected to isolate ligands showing high consensus sequence homologies to the known Src ligands (even if it was hoped to find ligands with higher binding affinities). It was assumed that having an idea about the expected type of sought ligands, would help characterizing the best experimental conditions, for the panning procedure against SH3 domains.

Src presents particular interests, since, along with the phosphatidylinositol 3-kinase (PI3K) SH3 domain, is the only SH3 domain to bind to ligands belonging to both classes (with RPLPPLP and XPPLPXR consensus sequences for Src, and RXLPPRP and KPPLPXR for PI3K, for class I and II respectively). However, for both Src and PI3K SH3 domains, class I ligands were predominantly isolated, with respect to the class II ligands (Yu et al., 1994; Sparks et al., 1996). Sparks et al. (1996) also reported that N terminal SH3 domain of Grb2

has the capacity to bind both classes I and II peptide ligands.

1-1-1 Libraries previously generated to define SH3 ligand preferences

The work done on Src SH3 domains, to determine the consensus sequence of their ligands, from phage display libraries, has been accomplished by many laboratories. However, three of them performed extensive work on Src SH3 domain. Their results are summarized in Table IV-2.

In 1994, Sparks et al. generated three random phage libraries which, after panning selection on Src, established that a sequence RPLPPLP is the preferred binding site for Src (Table IV-2). The first two libraries, were random peptide libraries displaying 22 or 36 amino acid long peptides, and containing $2x10^8$ unique variants. The third library has been established by taking advantage of the knowledge about Src, obtained from the analysis of Src natural ligands. They create a kind of biased library supposed to display peptides containing many of the residues identified in naturally occurring SH3-binding sequences. This library presents $((C/A)NN)_8$ peptides, and its diversity is $1x10^6$. This codon scheme encodes amino acids in the following proportions: 6R, 4P, 4L, 4T, 3I, 2H, 2Q, 2K, 2N, 2S, and 1M.

With this study, Sparks et al. showed that natural ligands and phage display isolated binders have similarities, and therefore, that it might be possible to compete *in vivo* the natural ligands, with synthetic peptides. In addition, they specified the Src consensus sequence (RPLPPLP), and proved that a 7 amino acid peptide is able to show binding capacity for the Src SH3 domain, whereby the presence of flanking regions appear to increase the affinity. Finally, by checking the relative binding of their selected peptides and potential cross reaction with other SH3 domains, they demonstrated that SH3 domains possess distinct ligand preferences.

In 1994, Rickles et al., also generated phage display libraries, in order to isolate ligands against Src. They first prepared a 6 residue peptide, flanked by glycine. After 8 cycles of panning, they sequenced 10 clones, which were all identical. Interestingly, a point mutation occurred, and the first glycine was replaced by an arginine, to give a RSLPPIPG sequence. These results confirm the importance of an arginine just prior the proline-rich motif. They assumed that the proline-rich motif was important for Src recognition, and they created a second biased library, consisting of 6 random amino acids, flanking the proline rich motif defined with the first library: X_6**PPIPG**. After three cycles of binding selection, the sequenced clones show the expected RPL consensus sequence before the proline fixed motif

(Table IV-2). To investigate the sequence preferences in the proline motif of the Src SH3 ligands, another biased library was generated: **RSLRPL**X_6. The C-terminal consensus sequence derived from this library was PPLPXP, a sequence containing the Src consensus sequence previously characterized by Sparks et al., RPLPPLP.

Also in 1994, Yu et al. designed and constructed a biased combinatorial peptide library of the form X_3**PP**X**P**X_2, containing $2x10^6$ variant peptides. The screen of this library on Src SH3 domain, allowed the characterization of ligands which can be categorized into two distinct classes, the class I and class II ligands, showing RXL**PPLP** and **PP**L**P**XR consensus sequences, respectively (Table IV-2). It has to be specified that the class I clones were largely dominant. Yu et al. suggested that these two classes of peptides bind to Src SH3 domain in reverse orientations, and confirmed it by structural analysis of SH3-ligand complexes (Feng et al., 1994).

Library	Consensus sequence	References
X_{22} X_{36} $((C/A)NN)_8$	**RPLPPLP**	Sparks et al., 1994
GX_6G X_6PPIPG RSLRPLX_6	***R*SLPPIP**G XXX**RPL**PPIPG RSLRPL**PPLP**X**P**	Rickles et al., 1994
X_3PPXPX_2	**R**XL**P**P**L**P P**P**L**P**X**R**	Yu et al., 1994
X_5RPLPPLP GAAPPLPPRX_5	X**SL**XX**R**PLPPLP GAAPPLPPR**N(R/K)**X**RL**	Rickles et al., 1995
X_6PXXPX_6	**L**XX**RPL**P**X**Ψ**P**X_6	Sparks et al., 1996

Table IV-2: Comparison of Src SH3 consensus sequences, obtained from ligands isolated with different phage display libraries (combinatorial peptide library for the work done by Yu et al., 1994). Underlined residues are fixed within the library. Red amino acids represent the determined ligand preferences of Src SH3 domain. The italic residue, in consensus motif obtained from GX_6G (Sparks et al., 1994) represents a point mutation which replaces a G by a R. Ψ stands for hydrophobic residues.

Then, in 1995, Rickles et al. developed new libraries, to further explore the ligand residues important for specific SH3 domain recognition. They stressed the investigation of the flanking regions other than the arginine responsible for the classification as class I or II. They

generated two banks, of the form X_5**RPLPPLP** and **GAAPPLPPR**X_5, to study additional Src SH3 binding preferences, of class I and class II ligands. After four rounds of selection, they isolated clones showing distinct consensus sequences, of the form SLXX**RPLPPLP** for the class I ligands, and **GAAPPLPPPR**N(R/K)XRL for class II (Table IV-2).

In 1996, Sparks et al. came back to a library which gave the possibilities to the SH3 domains to interact with a peptide showing more freedom, since only two prolines were fixed. Furthermore, this bank explored 6 residues on both sides of the proline core motif, allowing selection of both class I and II ligands. This bank was of the form X_6**P**XX**P**X_6, and its diversity was $3x10^8$ variants. However, the clones isolated from this bank, after panning against Src SH3 domain, were nearly exclusively belonging to the class I ligands, with the LXXRPL**P**XΨ**P** consensus sequence (Table IV-2), confirming the predominance of class I ligands, and the importance of the RPL residues, before the proline rich motif. Only one clone shows a sequence which might be related to the class II ligands. However, Sparks et al. did not investigate the relative binding capacity of this clone to the Src domain.

1-1-2 Advantages of the CPLPPXP library over previously described banks

The CPLPPXP library used in this work to isolate proline rich ligands against SH3 and EVH1 domains, looks similar to some of the banks shown above. However, although it was intended to characterize ligands showing high consensus sequence homologies to the ligands previously described, the aim of the work was to characterize ligands which show higher binding affinity for their target.

The main limitation of the banks utilized in the work already done, was their diversities (Table IV-3). The libraries generated by Sparks et al. (1994, 1996) contain a maximum of $3x10^8$ variants, for banks which consist of 14, 22, or 36 random amino acids. With $3x10^8$ variants within a library, one can expect to have only 5 or 6 positions containing all possible amino acids, with a 99% confidence (Table III-1). The library prepared by Yu et al. (1994) is even less complete. Its diversity was estimated to be $2x10^6$, however, a bank containing 6 random amino acids would require at least $2x10^9$ transformants (Tables III-1 and IV-3). The diversities of libraries synthesized by Rickles et al. (1994, 1995) were appropriate and should allowed, for most of the banks, the randomisation of all 5 or 6 random amino acids.

The CPLPPXP possessed a diversity of $2x10^8$ variants, which is similar to the sizes of the libraries developed by Sparks et al. (1994, 1996). However, the number of random amino

acids was reduced to 10 positions (compared to 14, 22, or 36). The libraries generated by Sparks et al. (1994), during the first part of their work, were appropriate to the knowledge of the moment about SH3 domains, but it seems nowadays more interesting to employ biased libraries, for a general work on a panel of SH3 domains. Their choice of a $((C/A)NN)_8$ bias has limitations, since only a fraction of the codons were represented in the library. This type of bank might in most of the cases prevents isolation of ligands for particular targets, since 9 of the amino acids are not represented. The biased library developed in 1996 by Sparks et al., was ingenious, as demonstrated by the resulting ligands, for a large variety of SH3 domains. A minimum proline motif was fixed, and a large freedom was given to both proline flanking regions (X_6PXXPX_6). However, this library suffered a serious diversity problem (Table IV-3), i.e. in the potentially diversified regions containing 14 amino acids, probably a maximum of 6 residues would be optimized. This might have limited the number of characterized ligands.

Library	Random positions	Diversity	Required diversity 90%	Required diversity 99%	References
X_{22}	22	$2x10^8$	$2x10^{33}$	$4x10^{33}$	Sparks et al., 1994
X_{36}	36	$2x10^8$	$3x10^{54}$	$6x10^{54}$	
$((C/A)NN)_8$	8	$1x10^6$	$2,5x10^{12}$	$5x10^{12}$	
GX_6G	6	$2,5x10^9$	$2,5x10^9$	$5x10^9$	Rickles et al., 1994
X_6PPIPG	6	$2x10^9$			
$RSLRPLX_6$	6	$5,8x10^8$			
X_3PPXPX_2	6	$2x10^6$	$2,5x10^9$	$5x10^9$	Yu et al., 1994
$X_5RPLPPLP$	5	$7,5x10^7$	$7,7x10^7$	$1,6x10^8$	Rickles et al., 1995
$GAAPPLPPRX_5$	5	$2,2x10^7$			
X_6PXXPX_6	14	$3x10^8$	$2,5x10^{21}$	$5x10^{21}$	Sparks et al., 1996

Table IV-3: Comparison of estimated and required diversities, of combinatorial and phage display libraries, generated by other groups to isolate ligands against SH3 domains. Required diversities represent the necessary sizes of the banks, according the their number of random amino acids, to ensure representation of all possible peptides, with a 90% or 99% confidence level. The confidence that a library contains all possible amino acid sequences is calculated assuming a Poisson distribution (Lowman and Wells, 1991).

The CPLPPXP library is similar to the bank of Yu et al. (X_3PPXPX_2; 1994), since in both cases, a Pro Pro Xxx Pro core motif was introduced within the hypervariable region. However, CPLPPXP diversity was significantly higher that of Yu et al. In addition, the library

used by Yu et al. (1994) was a combinatorial chemistry library, where isolation of ligands is limited to a single round of selection. There, one does not analyze single sequences, but obtains an impression of a consensus sequence. If there are several possible sets of consensus sequences, information is lost with this method.

The sizes of the proline motif flanking regions were more appropriate in the case of the CPLPPXP library, since in the Yu bank, the randomized N and C terminal parts of the proline motif, only allowed isolation of ligands showing class I or II characteristics. The flanking regions of SH3 ligands are responsible for specificity between the ligand, and a particular SH3 domain (the proline rich motif is common to every SH3 domain). However, since the SH3 domains are classified in two groups, according to their ability to interact with ligands from class I or II, the RXL (from **RXL**PPXP class I ligand), and XR (from PPLP**XR** of class II ligands) are still assumed to represent a motif common to a group of SH3 domains. Therefore, to really learn about SH3 domain specificity, longer flanking regions are required.

The libraries designed by Rickles et al. (1994, 1995) allowed characterization of ligands against a large variety of SH3 domains. The libraries had an adequate diversity, and were so successful due to step by step successive optimization, termed the “extension library” approach. Their strategy was to use a random library against SH3 domains, to search for consensus motif against a particular SH3 domain. Once obtained, they created a secondary library, presenting the defined motif, and additional random amino acids, to further explore the flanking regions and extend the determined primary motif. The next library was then applied, obtaining larger consensus sequences. The only drawback of their work, is the short size of the random peptides displayed on the surface of the phage (6 residues), used in the primary library, compared to the number of fixed positions (5 residues) within the secondary bank. Some specific ligands which might have made use of novel synergistic interaction between sub-segments may have been lost by defining so many positions.

Despite the limited diversity of the CPLPPXP library, the bank was considered to be large enough, since it was designed as a cosmix-plexing® library, and therefore, can be recombined as required, increasing the size of the bank.

However, the main advantage of the cosmix-plexing®, is the recombination among peptides of a preselected population of clones which have at least weak affinity for the target (i.e. enriched for “target significant” sequence space). The populations selected after one or two rounds of panning selection were collected, and submitted to cosmix-plexing®

recombination. This leads to the reassortment of the amino and carboxy proximal sections within the hypervariable domain, among clones which are highly enriched in SH3 ligands.

Another reason to believe that the CPLPPXP library was particularly suitable for the isolation of high affinity SH3 ligands, was that CPLPPXP is a **phagemid** library. Sparks et al. (1994, 1996), and Rickles et al. (1994, 1995), worked with **phage** display libraries, by fusing their random peptides to the pIII phage protein. In the CPLPPXP library, the pIII protein is also used as fusion partner, however, a phagemid system was used. In phage libraries, the random peptides are present five times on the surface of the phage, while in the CPLPPXP phagemid library, they were expected to be essentially displayed once per particle. Monovalent presentation is of high advantage when strong affinity ligands have to be isolated. The reduced valency is valuable in eliminating avidity. Only the latter method permits real discriminative selection for high affinity as compared to modest affinity peptides.

1-1-3 Src ligands, isolated from CPLPPXP library

The panning procedure on Src SH3 domain was performed in parallel under various conditions, to determine the optimal affinity selection protocol for SH3 domains. Microtiter plates were used as support material for the panning. It has been shown that glutathione microtiter plates were not applicable to the screening on GST-Src fusion protein. Such a plate was supposed to capture the GST part of the GST-SH3 fusion protein, with the glutathione fixed on the wells, and therefore, the SH3 protein was expected to be orientated such that it was available for interactions with potential ligands. Surprisingly, using glutathione plates, no enrichment was observed, and the output values were very similar to the one of the blank, suggesting that no selection occurred. The results indicated that the GST-Src fusion proteins were possibly not retained sufficiently tightly by the glutathione, or that the interaction between the glutathione and the GST might have oriented the Src molecules in manner unfavorable for interactions with their ligands. Therefore, maxisorb microtiter plates were employed, and provided satisfactory results.

The best washing conditions were defined experimentally, aiming to achieve an equilibrium between washing away weak binders and thus, removing the background, but still retaining interesting ligands. This is of particular importance during the first cycle of selection, where discrimination between non-specific, and specific binders takes place within the starting population, where the specific ligands may be still under-represented. This was obtained by using mild washing steps, with a PBS buffer containing 0,05% of Tween 20. The stringency was increased with each cycle of selection, since after each round of panning, the

eluted phage were amplified. This means that the high affinity ligands are more and more enriched, and higher stringency is required for better discrimination.

The elution of the bound phage was achieved by using a low pH glycine buffer.

The panning procedures were carried out in the same way for all other targets.

The screening of the CPLPPXP library on the Src SH3 domain led to the isolation of ligands which allowed the conception of a clear consensus sequence (Figure IV-2). In the case of the Src domain, cosmix-plexing® recombination of the population selected after first and second rounds of panning led to the enrichment, directly after the first round, of a unique clone (S1), identical to the clone which already dominated over the other ligands, after standard panning. This demonstrates a very strong binding preferences of Src SH3 domain, for this S1 clone, since reassortments amongst pre-selected clones did not allow the creation of new clones which would have shown new features, optimal for the target. The clones selected for Src SH3 domain, are presented in the Figure IV-2.

	Sequences against Src SH3 domain														**E**	**B**
	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}	P_{13}			
S1	Gly	Asn	**Arg**	**Pro**	**Leu**	**Pro**	**Pro**	**Ile**	**Pro**	Ser	His	**Pro**	Phe		+	+
S4	Lys	Gly	**Arg**	**Pro**	**Leu**	**Pro**	**Pro**	**Val**	**Pro**	Gly	Thr	**Pro**	**Ser**		+	+
S2			**Arg**	**Pro**	**Leu**	**Pro**	**Leu**	**Pro**	**Pro**	**Val**	**Pro**	Trp	Val	Arg Trp	+	+
S3			**Arg**	**Ser**	**Leu**	**Pro**	**Leu**	**Pro**	**Pro**	**Val**	**Pro**	Thr	**Ser**	Ser Ala	+	-
Consensus sequence																
	Xxx	Xxx	**Arg**	**Pro**	**Leu**	**Pro**	**P/L**	**Ψ**	**Pro**	**Ψ**	**Pro**	**Pro**	**Ser**			
S5	Ser	His	Tyr	**Ser**	**Leu**	**Pro**	**Pro**	Gly	**Pro**	**Leu**	Arg	Gly	Phe		-	-
S6	Asn	Ala	Cys	Thr	**Leu**	**Pro**	**Pro**	**Phe**	**Pro**	**Trp**	Asp	Arg	Cys		-	-
S7	Gly	Ala	Gly	**Ser**	**Leu**	**Pro**	**Pro**	Cys	**Pro**	**Ala**	Asp	Leu	Leu		-	-

Figure IV-2: Sequences of single clones isolated against Src SH3 domain, and the deduced consensus sequence. The clones are classified according to their relative affinity on phage ELISA, for the target. Only the clones positive in phage ELISA and/or BIAcore measurements, are taken into account for the conception of the consensus sequence. A residue is considered as representative for the consensus sequence, when it occurs in 50% or more of the sequences. Underlined amino acids are fixed residues within the CPLPPXP library. Ψ represents hydrophobic amino acids. E and B stand for phage ELISA and BIAcore analysis. A simple + or – indicates whether the clones reacted positively or not, in ELISA or BIAcore tests. P_1 to P_{13} refer to the amino acid position within the sequence. For occurrence of each clone, or precise phage ELISA values, refer to Figures III-21 and III-23.

Only four of the seven enriched ligands are shown to react strongly *in vitro* with the Src SH3 domain. However, these four clones allowed the obtention of a long consensus sequence

(11 amino acids). These four ligands are all showing characteristics belonging to the class I ligands, mainly with the presence of a conserved arginine, at P_3, before the proline motif, but also with the Arg Pro Leu motif. Src SH3 domain has been previously shown to be able to interact with ligands (natural or not), belonging to both classes (Yu et al., 1994; Feng et al., 1994; Sparks et al., 1996). However, class I ligands largely dominate over class II (Figures IV-3 and IV-4).

It should be noticed that a clone which reacts positively on phage ELISA does not necessarily interact with the target under the conditions of the BIAcore analysis. One example is clone S3, positive in ELISA (even if it was the weakest positive clone), but negative with Src in the BIAcore test. This might be explained by the absence of Pro at position P_4, which has been demonstrated to be required for binding of the ligands to Src SH3 domain. However, it is not expected to have always similar reactions between phage ELISA and BIAcore. Indeed, in BIAcore, the ligands are in flow. This means that if the dissociation is fast, the ligands are released from the target, and washed away. In phage ELISA, when dissociation between both partners is fast, the ligands are still present in the microtiter wells, and still have a chance to interact with the target, and the ELISA signal, in such a case, might be stronger than observed on BIAcore. In phage ELISA, there is little discrimination between low or high dissociation rate.

The S5 clone is the only selected clone which might be considered as a class II ligand, due to the arginine at P_{11}. However, S5 does not show leucine at P_8, as class II ligands for Src SH3 domain do. This is may be why the S5 clone is not reacting on phage ELISA.

The S1 clone was the clone isolated most frequently in standard panning, and the sole clone selected after cosmix-plexing® recombination. In phage ELISA, it shows the best relative affinity for the Src domain. S1 might be so advantageous for the Src domain, because it could be classified under both classes of ligands. The Arg Pro Leu motif, at P_3-P_5, are characteristics of class I ligands. However, the His at P_{11}, and the Ile at P_8 suggest a class II ligand. Class II ligands usually present an Arg at P_{11}, and not the His found in S1. However, His and Arg, are both basic amino acids, and the histidine might be able to establish the required stabilizing salt bridge with the conserved Asp within the Src SH3 domain. Such clones have already been reported by Rickles et al. (1995), against the PI3K SH3 domain: a phage displaying RSL<u>RPL</u>**<u>PPLP</u>PR**PXX was selected, which also presents both class I (underlined) and class II (boldface type) consensus sequences. Such clones might potentially bind in either orientation to the SH3 domain.

The S1 displayed peptide has been synthesized as free peptide, and was used as competitor in competitive phage ELISA. Src SH3 domain was immobilized in a microtiter plate well, and incubated with fixed concentrations of S1, S2, S3, or S4 phage variants. After washing away non bound phage, increasing concentrations of the synthetic free S1 peptide was added. The amount of bound phage remaining was estimated, after new washing steps (Figure III-44). All four clones were competed away from the Src binding site, by the free S1 peptide. This demonstrates, as it was expected, that all four peptides were interacting with Src SH3 domain, at the same binding site. But moreover, it proves that the linear S1 displayed peptide properties, were transferable between phage-bound peptide, and phage-free peptide. This is not the case for every phage displayed peptide, since the linear peptide might be constrained by its local environment (Ladner, 1995). This feature promotes the S1 peptide as valid for *in vivo* testing.

1-1-4 Comparison of Src ligands, from various libraries

The S1 clone sequence was compared with Src ligands isolated by other groups (Sparks et al., 1996: X_6PXXPX_6; Yu et al., 1994: X_3PPXPX_2), with biased library, displaying a proline core motif. The Figure IV-3 presents the sequences.

The comparison of the class I Src ligands, obtained from different biased libraries (Figure IV-3), shows distinct features. This was expected, since the three libraries used for the selection of binders studied here, were based on various fixed core sequences in addition to differences in the valency of presentation. The CPLPPXP library is a phagemid library. Sparks et al., used a phage library, and Yu et al., employed a combinatorial chemistry library.

The common characteristic of the ligands isolated from these three libraries, is the motif before the proline core. Arg at P_3 and Leu at P_5 appear to be essential for the ligands. However, the Pro at P_4, which was demonstrated as crucial, in our experiments, for high affinity binding between the ligands and the Src domain (ligands without Pro at P_4, lost their binding for Src on phage ELISA; see Figure III-31), is present only in half of the clones of Sparks, and in 1/4 of the variants of Yu.

However, it has to be emphasized that the CPLPPXP library was designed to isolate strong ligands against SH3 domains, while the phage library of Sparks, with a display of 5 peptides per particle, created an avidity phenomenon, and therefore, tended to isolate weaker affinity binders. Furthermore, their sequenced clones were picked from the third round of

panning, before a strong enrichment of the higher affinity ligands occurred, and most of their clones were isolated only once (the sequenced clones, previous to sequencing, were shown to react positively on phage ELISA). As far as the library of Yu was concerned, since it was a combinatorial chemistry library, only one round of selection was possible, and no significant discrimination between low or high affinity ligands was possible. So, the fact that Pro at P_4 is not systematic in Sparks and Yu libraries, might simply reflect isolation of weaker binders.

Major differences appear on the flanking region C terminal from the proline motif (Figure IV-3). In our consensus sequence, a stretch of proline or hydrophobic residues appeared to be advantageous for the interaction (Figure III-23).

For Sparks clones, no clear consensus can be conceived from this region, however, a significantly high amount of acidic residues (Asp and Glu) is found, at various positions, between P_{10} and P_{14}, since 8 from the 13 isolated clones present at least one acidic amino acid at one of these positions. A comparison with the ligands isolated with the CPLPPXP library demonstrate that only ligands which are too weak to show a positive interaction with the Src domain on phage ELISA, also possess an Asp (however, localized at P_{11}, and present in 2/3 of the clones; Figure IV-2).

Similarly, in Sparks clones, several prolines are found, also between P_{10} and P_{14} (8 from the 13 isolated clones possess at least one proline at one of these positions).

Curiously, an association between Pro and an aromatic residue (preferentially Phe) is observed several times, within the C terminal flanking region, in both ligands from CPLPPXP (S1 and S2) and Sparks libraries (overlaid in yellow).

For Yu ligands, a clear consensus sequence is observed on the C terminal flanking region, even though limited to only two residues. Arg, or eventually other basic residues, are clearly defined at P_{10}, while aromatic residues determined the position P_{11}.

Figure IV-3: Comparison of Src SH3 domain class I ligands, isolated from various biased libraries. The S1 clone was isolated from the CPLPPXP phagemid library. Sparks sequences were selected from a X_6PXXPX_6 phage library (Sparks et al., 1996). Yu sequences are derived from a X_3PPXPX_2 combinatorial chemistry library (Yu et al., 1994). Consensus sequences for Sparks and Yu clones are given. A residue is considered as representative for the consensus sequence, when it occurs in 50% or more of the sequences. Ψ and φ represent hydrophobic and aromatic amino acids, respectively. Underlined amino acids are fixed residues within the libraries. P_1 to P_{13} refer to the amino acid position within the sequence.

Sequences against Src SH3 domain – Class I ligands

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}	P_{13}
S1 dominant clone (X_5PPXPX_4)	Gly	**Asn**	**Arg**	**Pro**	Leu	**Pro**	**Pro**	**Ile**	**Pro**	Ser	His	**Pro**	Phe
X_5PPXPX_4 consensus sequence	Xxx	Xxx	**Arg**	**Pro**	Leu	**Pro**	**P/L**	**Ψ**	**Pro**	**Val**	**Pro**	**Pro**	**S/Ψ**

Sparks sequences (X_6PXXPX_6)

		Leu	Ala	Ser	**Arg**	**Pro**	Leu	**Pro**	Leu	Leu	**Pro**	Asn	Ser	Ala	Pro	Gly	Gln
		Leu	Thr	Gly	**Arg**	**Pro**	Leu	**Pro**	**Ala**	Leu	**Pro**	Pro	Pro	Phe	Ser	Asp	Phe
		Pro	Ala	Tyr	**Arg**	**Pro**	Leu	**Pro**	Arg	Leu	**Pro**	Asp	Leu	Ser	Val	Ile	Tyr
Arg	Ala	Leu	Arg	Val	**Arg**	**Pro**	Leu	Pro	**Pro**	**Val**	Pro	Gly	Thr	Ser	Leu		
Gly	Ser	Leu	Pro	Phe	**Arg**	**Pro**	Leu	Pro	**Pro**	**Val**	Pro	Thr					
		Leu	Lys	Val	**Arg**	**Ala**	Leu	**Pro**	Pro	Leu	**Pro**	Glu	Thr	Asp	Thr	Pro	Tyr
		Ile	Ser	Gln	**Arg**	**Ala**	Leu	**Pro**	Pro	Leu	**Pro**	Leu	Met	Ser	Asp	Pro	Ala
		Leu	Thr	Ser	**Arg**	**Pro**	Leu	**Pro**	Asp	**Ile**	**Pro**	Val	Arg	Pro	Ser	Lys	Ser
		Asn	Thr	**Asn**	**Arg**	**Pro**	Leu	**Pro**	Pro	Thr	**Pro**	Asp	Gly	Leu	Asp	Val	Arg
		Met	Lys	Asp	**Arg**	**Val**	Leu	**Pro**	Pro	**Ile**	**Pro**	Thr	Val	Glu	Ser	Ala	Val
		Leu	Gln	Ser	**Arg**	**Pro**	Leu	**Pro**	Leu	**Pro**	**Pro**	Gln	Ser	Ser	Tyr	Pro	Ile
		Phe	Ile	**Asn**	**Arg**	Arg	Leu	**Pro**	**Ala**	Leu	**Pro**	Pro	Asp	Asn	Ser	Leu	Leu
				Phe	**Arg**	**Ala**	Leu	**Pro**	Leu	**Pro**	**Pro**	Thr	Pro	Asp	Asn	Pro	Phe
Sparks consensus sequence		Leu	Xxx	Xxx	**Arg**	**Pro**	Leu	**Pro**	**Ψ**	**Ψ**	**Pro**	Xxx	Xxx	Xxx	Xxx	Xxx	Xxx

Yu sequences (X_3PPXPX_2)

Arg	Ala	Leu	**Pro**	**Pro**	Leu	**Pro**	**Arg**	Ala
Arg	Glu	Leu	**Pro**	**Pro**	Leu	**Pro**	**Arg**	**Phe**
Arg	Ala	Leu	**Pro**	**Pro**	Leu	**Pro**	**Arg**	**Tyr**
Arg	Asn	Leu	**Pro**	**Pro**	Leu	**Pro**	**Arg**	Ile
Arg	Ala	Leu	**Pro**	**Pro**	Leu	**Pro**	**Arg**	**Trp**
Arg	Thr	Leu	**Pro**	**Pro**	Leu	**Pro**	**Arg**	**Phe**
Arg	Ala	Leu	**Pro**	**Pro**	Leu	**Pro**	**Arg**	Ile
Arg	**Pro**	Leu	**Pro**	**Pro**	Leu	**Pro**	Thr	Ser
Arg	Met	Leu	**Pro**	**Pro**	Leu	**Pro**	Ala	**Trp**
Arg	**Pro**	Leu	**Pro**	**Pro**	Leu	**Pro**	Ala	**Trp**
Arg	Gln	Leu	**Pro**	**Pro**	Leu	**Pro**	Ala	**Phe**
Arg	Arg	Leu	**Pro**	**Pro**	Leu	**Pro**	Gln	Leu
Arg	Asp	Leu	**Pro**	**Pro**	Leu	**Pro**	**His**	Arg
Arg	Glu	Leu	**Pro**	**Pro**	**Ile**	**Pro**	Val	**Phe**
Arg	**Pro**	Leu	**Pro**	**Pro**	**Val**	**Pro**	Met	**Phe**
Arg	**Pro**	Leu	**Pro**	**Pro**	Thr	**Pro**	**Lys**	**Tyr**

Yu consensus sequence

Arg	Xxx	Leu	**Pro**	**Pro**	Leu	**Pro**	**Arg**	**φ**

Determination of the K_D of the ligands was performed. Sparks et al. (1996) estimated the dissociation constant of their ligands, by comparing the phage ELISA signals of their positive ligands to those obtained with previously characterized Src SH3 binding clones, and the K_D were assumed to range between 5 and 75 µM.

Yu et al. (1994), evaluated the K_D of some of their clones, by fluorescence titration assays, and found values, between 7 and 70 µM. In both cases, determined K_D values are similar to those for natural SH3 ligands, however ligands with a K_D of 0,2 µM were previously characterized, providing evidence that the ligands isolated by Sparks et al. (1996), and by Yu et al. (1994) are not amongst those with the highest affinities.

The K_D of the clones isolated from the CPLPPXP library were determined by BIAcore real time interaction measurements, and a K_D value of 7,3 nM was obtained for the S1 clone (Figure III-50), around 1000 fold better than the best clones isolated by Sparks et al. (1996) or by Yu et al. (1994). Actually, S1 was the only Src ligand for which the K_D value was properly defined, since the Src SH3 domain stock used for these tests were not concentrated enough to repeat the experiment performed with Src, and to reach the equilibrium phase necessary for the evaluation of the K_D. However, they are expected to show similar range dissociation constants, since their plots of Req versus Src concentration, looked very similar to that of S1 (Figure III-51).

The Table IV-4 presents ligands previously characterized against diverse SH3 domains. According to the huge number of existing ligands, whether natural or isolated from phage display techniques, a severe selection was done, to present only the highest affinity SH3 binders.

Until now, the highest affinity SH3 natural ligands was the Nef protein, from HIV-1, which binds with a K_D of 0,25 µM to the Hck SH3 domain (belongs to the Src SH3 family). The highest affinity ligands characterized for Src, where isolated by Cheadle et al. (1994), and Rickles et al.(1995), and are in the range of 0,2 µM.

The S1 Src ligand, isolated with the cosmix-plexing® CPLPPXP library, has a K_D of 7,3 nM. To our knowledge, this is the first time that a SH3 ligand possessing a K_D in the range of nM is described. The affinity of S1 for Src is around 30 fold higher than the best characterized ligands until now (from Cheadle and Rickles), however, it is much higher than natural ligands (around 3500 fold higher than SOS natural ligand) and might be able to compete *in vivo* experiment (considering work with cells non infected by HIV).

References	Natural ligands	SH3	K_D
Viguera et al., 94	3BP-1	Abl	34 µM
	3BP-1	Fyn	34 µM
	3BP-1	Src	200 µM
	3BP-2	Abl	5 µM
	SOS1-3	Src	27 µM
	SOS1-4	Src	26 µM
Lee et al., 95	Nef (HIV-1)	Hck	0,25 µM
	Nef (HIV-1)	Fyn	0,38 µM

References	Optimized natural ligands	SH3	K_D
Pisabarro et al., 96	3BP-1	Abl	34 µM
	opt 3BP-1	Abl	0,4 µM

References	Combinatorial library clones	SH3	K_D
Yu et al., 94	RKLPPRPSK	PI3K	9,1 µM
	RALPPLPRY	Src	7,8 µM
Cheadle et al., 94	TGPRPLPLPPLRSMS	Src	0,18 µM
Rickles et al., 94	PPPYPPPIP	Abl	2,0 µM
	RSSRPLPPIP	Src	19,5 µM
Rickles et al., 95	RPLPPLPGGK	Src	17,7 µM
	VSLARRPLPPLPGGK	Src	0,86 µM
	KGGGAAPPLPPR	Src	20,3 µM
	KGGGAAPPLPPRNRPRL	Src	0,24 µM
Horn and Collins	GNRPLPPIPSHPF (S1clone)	Src	**7,36 nM**
(in preparation)	KGRPLPPVPGTPS (S4 clone)	Nephrocystin	**0,42 nM**

Table IV-4: K_D values of some SH3 ligands. Ligands were selected according to their K_D values, only the higher affinity ligands are shown. Presented SH3 ligands were isolated by different groups, and are either original natural ligands, optimized natural ligands, or ligands characterized by combinatorial libraries (phage display libraries, except Yu et al., (1994) who worked with combinatorial chemistry library). The highest affinity ligands characterized until now are shown in blue (µM range), the S1, and S4 clones identified against Src and nephrocystin SH3 domains from CPLPPXP library are presented in red (nM range).

As far as the class II ligands are concerned, not a lot can be said, according to the low number of isolated ligands (Figure IV-4).

From the two ligands isolated from CPLPPXP library, S1 cannot completely be considered as a class II ligand, since it does not show an Arg, but a His at P_{11}. The S5 ligand does not react on phage ELISA with Src SH3 domain. The unique clone described from Sparks variants has not been demonstrated to bind to the target *in vitro*. And only one of the

clones of Yu et al. interacts with Src, however, with a low K_D (73 µM). Therefore, no real consensus sequence can be designed.

Sequences against Src SH3 domain – Class II ligands															
CPLPPXP library	**(X_5PPXPX_4)**														
S1		Gly	Asn	Arg	Pro	Leu	Pro	Pro	Ile	Pro	Ser	His	Pro	Phe	
S5		Ser	His	Tyr	Ser	Leu	Pro	Pro	Gly	Pro	Leu	Arg	Gly	Phe	
Sparks sequences	**(X_6PXXPX_6)**														
	Leu	Tyr	Ser	Ala	Ile	Ala	Pro	Asp	Pro	Pro	Pro	Arg	Asn	Ser	Ser	Ser
Yu sequences	**(X_3PPXPX_2)**														
				Arg	Asp	Leu	Pro	Pro	Leu	Pro	His	Arg			
				Pro	Tyr	His	Pro	Pro	Leu	Pro	Arg	Arg			
				Ala	Phe	Ala	Pro	Pro	Leu	Pro	Arg	Arg			
Similarities between the sequences															
				Xxx	Xxx	Ψ	Pro	Pro	Ψ	Pro	β	Arg			

Figure IV-4: Comparison of Src SH3 domain class II ligands, isolated from various biased libraries. The S1 clone was isolated from the CPLPPXP phagemid library. Sparks sequences were selected from a X_6PXXPX_6 phage library (Sparks et al., 1996). Yu sequences are derived from a X_3PPXPX_2 combinatorial chemistry library (Yu et al., 1994). Underlined amino acids are fixed residues within the libraries. Ψ and β represent hydrophobic and basic amino acids, respectively. P_1 to P_{13} refer to the amino acid position within the sequence.

1-1-5 Swissprot comparison of Src ligands isolated from CPLPPXP library

A computer-assisted sequence search with the Swissprot database failed to identify existing proteins showing homology with the Src isolated ligands, except for the S3 clone. This is not surprising that the high affinity isolated ligands do not resemble natural proteins. Our ligands are shown to have extremely high K_D for the Src SH3 protein. However, natural ligands require a weaker interaction, in order to allow sensitive and dynamic modulations in response to changing signals. That might be the reason why the only clone (S3) for which homologue proteins were observed, was the clone showing weaker relative affinity in phage ELISA, and was not positive in BIAcore measurements.

The Swissprot database revealed that the S3 sequence shows 70% identity in a stretch of

13 amino acids with the human WASP (Wiskott-Aldrich Syndrome Protein), and 89% identity in a stretch of 9 residues, with the rat CAP1 protein (Adenylyl cyclase associated protein 1), as shown in Table IV-5.

S3 ligand Human WASP	**RS**L**PLPPVP**T *S*S**A** **RS**G**PLPPVP**L*G*I**A**	}	70% identity in 13 aa overlap
S3 ligand Rat CAP1	RSL**P**L**PPVPTSS***A* PGP**P**P**PPVPTSS***G*	}	89% identity in 9 aa overlap

Table IV-5: Amino acid of S3 clone and comparison to human WASP (amino acids 338-350), and rat CAP1 (amino acid 237-245), according to Swissprot database analysis. Regions of identity are shown in red, and of homology in blue.

1-1-5-1 CAP1 homologies

The identity between the S3 clone and the CAP1 was first thought unlikely to be of significance, since the first Arg, along with the following Leu, at the third position, are assumed to be crucial for the interaction with Src, but these three amino acids do not belong to the region which has similarity with the CAP1 sequence. This reduces the chance that CAP1 might be a natural ligand of Src SH3 domain.

However, the yeast CAP protein has been proved to be involved in interaction with the SH3 domain of the Abp1p yeast protein (Freeman et al., 1996). Genetic studies suggested two roles for CAP which are independently regulated. One as a positive regulator of cAMP levels in yeast (Shima et al., 1997), and a second as a cytoskeletal regulator, through an actin sequestering activity (Freeman et al., 1995). CAP protein contains a proline rich region subdivided into two domains (Yu et al., 1999). The first proline rich region contains a 10-12 amino acid stretch composed almost entirely of prolines. The second proline rich region contains a SH3 binding motif, which binds *in vitro* to SH3 domains, and is required to direct the CAP to corticol actin patches (Freeman et al., 1996). Recently, Yu et al. (1999) demonstrated that monomeric CAP efficiently binds actin and adenylyl cyclase, but not the Abp1-SH3. The formation of a dimeric CAP complex exposes the CAP SH3 binding site. In theory, CAP could serve as an adapter protein to direct adenylyl cyclase to actin corticol patches (Yu et al., 1999).

CAP1, the first characterized mammalian homologue of the yeast CAP protein, was isolated as actin-monomer binding protein (Gieselmann and Mann, 1992). All CAP homologues contain a unique centrally located proline-rich region, corresponding to the first

yeast proline rich region (10-12 amino acid stretch composed almost entirely of prolines).

The identity between the S3 clone, and the CAP1 concerned the CAP1 amino acids 237 to 245, which overlaps with the proline rich region of the rat CAP1, extended from amino acids 229 to 240. The data suggest that the Src SH3 domain might be a mammalian homologue of the yeast Abp1 SH3 domain, and that the S3 clone might mimic the CAP1 protein, for interaction with Src. However, if the mammalian and yeast CAP proteins have similar functions, interaction between S3 and Src should disrupted Src roles in actin organization. Indeed, since CAP has distinct binding sites for SH3 domain, actin monomer, and the adenylyl cyclase, but that the S3 clone would only provide a SH3 binding site, the adapter function would then be lost, and the adenylyl cyclase could not be directed anymore to the actin corticol patches. S3 might be a CAP1 competitor.

1-1-5-2 WASP homologies

The second protein which shows identity with the S3 clone, is WASP. The S3 region involved in the homology, is the same as the one expected to bind to Src SH3 domain, through the proline motif, completed with the Arg Xxx Leu motif before the proline core.

WASP has been shown to interact with the Itk, Src, Lck, Abl SH3 domains (Bunnell et al., 1996), the C terminal SH3 domains of Nck (Nck3) (Rivero-Lezcano et al., 1994; Anton et al., 1998), and with the N terminal Grb2 SH3 domain (She et al., 1997). Furthermore, The Itk, Src, Abl and Lck have been demonstrated to bind to distinct sets of proline rich regions in WASP (Bunnell et al., 1996).

Wiskott-Aldrich syndrome, the disease associated with defects in WASP is characterized by platelet and lymphocyte cell abnormalities, most likely caused by defects in actin cytoskeleton organization (Ochs et al., 1980). WASP is a proline rich protein (it contains 7 distinct proline rich regions) that has been shown to directly interact with activated Cdc42 (small GTP-binding protein; Symons et al., 1996), and with WIP (WASP Interacting Protein; Anton et al., 1998), which are both involved in actin cytoskeletal organization. Src is found ubiquitously, but most abundantly in platelets, brain and osteoclasts. The fact that WASP defects are associated with severe platelet anomalies, might be correlated with an obstructed interaction with Src (taking into account that Src SH3 domain is one of the binding partners of WASP).

Bunnell et al. (1996) and Anton et al., (1998) showed that interaction of WASP with Src, Abl, and Nck3 SH3 domains requires the presence of several WASP domains. Interaction

between Src and WASP requires the presence of two WASP fragments. Both sites have not been precisely mapped, so that it cannot be discriminated if both sites synergize to yield high affinity binding, or if interaction takes place through divalent binding. For the Itk SH3 domain, two domains are also required, but since they are consecutive, it might be that the binding site is split between both regions, and lies at their junction. However, the possibility that binding needs the presence of two binding sites cannot be ruled out.

The binding of Nck3 with the WASP protein involved a WASP region repeated twice, showing the consensus sequence GRSGPXPPXP (Finan et al., 1996). The S3 homology with WASP is concerning one of these regions since out of the 10 amino acids of the WASP stretch, 8 are identical to the S3 ligand.

S3 ligand **RS** L**PLPPVP**T*S*S**A**
Human WASP G**RS**G**PLPPVP**L*G*I**A**
Nck3 binding domain

This strongly suggests that S3 ligand interacts with Src in the same way that WASP and Nck3 are interacting (Figure IV-5/A). Additionally to the binding of WASP to the C terminal SH3 domain of Nck (Nck3), the WIP protein (WASP Interacting Protein) specifically interacts with its central SH3 domain (Nck2). WIP also specifically binds to WASP, to profilin (probably with two profilin molecules, since WIP has two profilin binding sites) and contains an actin binding site (Anton et al., 1998). The actin regulatory protein profilin binds to actin monomer. Therefore, WASP and WIP may link Nck to the actin cytoskeleton, and WIP might amplify the recruitment of actin monomers. Nck may simultaneously engage WASP and WIP, thereby increasing the local concentration of both proteins, and enhancing their interaction. Since Nck is recruited via its SH2 domain to phosphotyrosine residues in receptor tyrosine kinases following their activation by ligand binding, the Nck-WASP-WIP interaction may provide an important link between extracellular signaling via receptor tyrosine kinases and reorganization of the cytoskeleton.

Src also interacts via its SH2 domain with phosphotyrosine residues in receptor tyrosine kinases, and therefore, the homology between S3 and WASP suggests that Src could also be linked to the actin cytoskeleton through the WASP-WIP-profilin-actin interaction, as illustrated in Figure IV-5/B).

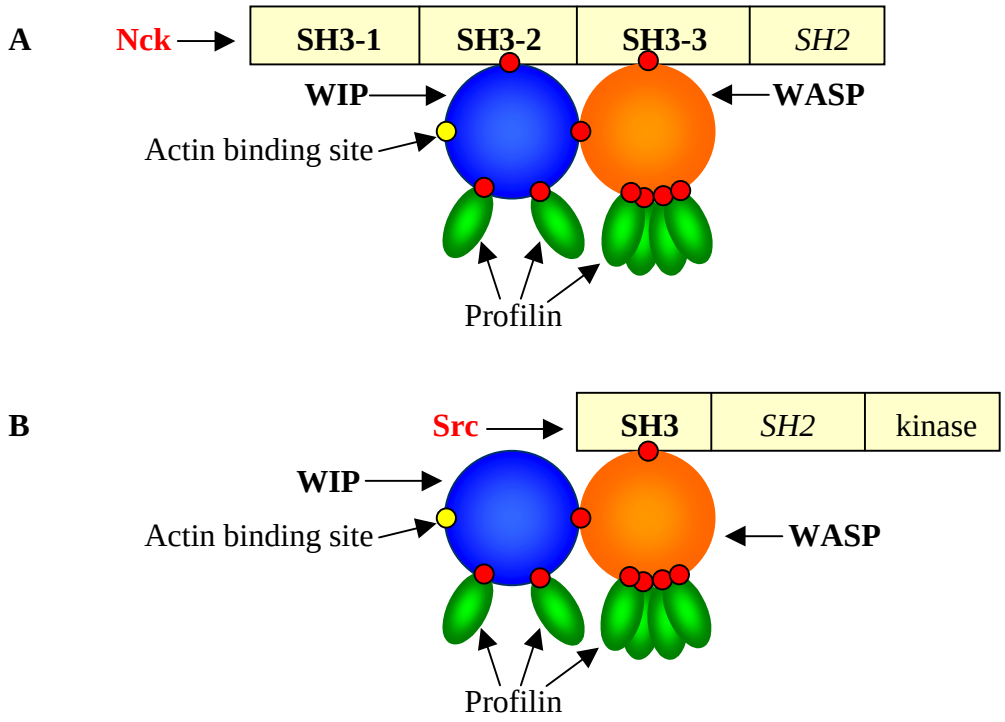


Figure IV-5: Illustration of Nck and potentially Src SH3 domains are linked to the actin cytoskeleton organization. (A-) WIP interactions with Nck, WASP and profilin as demonstrated by Anton et al. (1996). (B-) Hypothetical link between Src and actin organization. The red circles represent specific interaction between partner proteins.

1-2 Nephrocystin SH3 domain

The nephrocystin SH3 domain has been recently discovered by Hildebrandt's group (Hildebrandt et al., 1997a; Hildebrandt et al., 1998), and to our knowledge, was used here, for the first time as target in phage display, to isolate specific nephrocystin ligands.

Nephrocystin is a region of the *NPHP1* gene, and might be involved in juvenile nephronophthisis disorder, since 80% of the families suffering from this disease show a deletion within the nephrocystin gene. The nephrocystin region has been demonstrated to have strong homology and to be related, to previously characterized SH3 domains. The potential function of nephrocystin SH3 domain is still unknown. Therefore, characterization of nephrocystin SH3 domain ligands is of medical interest.

1-2-1 Nephrocystin ligands isolated from CPLPPXP library

The panning procedure was performed similarly to that for Src SH3 domain. However, the analysis of the output/input values, indicated enrichment factors very different from those observed for Src. After standard panning selection on Src, an enrichment factor of 120 was

obtained after the third round, and of 1500 after the fourth cycle, whereas for nephrocystin, the enrichment factor is inferior to 3 after three rounds of panning. However, a 20 fold enrichment was observed between the second and third round of affinity selection, and sequencing of randomly picked clones provided sequences of only three different clones (Figure IV-6), each recovered up to 7 times. Alignment of the three variants allowed the conception of a clear consensus sequence, which showed strong homologies with the consensus sequence defined for the Src SH3 domain, as shown:

Nephrocystin	**Arg**	Xxx	**Leu**	**Pro**	**Pro**	Xxx	**Pro**	**Val**	**Pro**	
Src	**Arg**	**Pro**	**Leu**	**Pro**	**P/L**	**Ψ**	**Pro**	**Ψ**	**Ψ**	**Pro**	**S/Ψ**

The population recovered after the first round of panning was submitted to cosmix-plexing® recombination, and the resulting secondary library was screened against nephrocystin SH3 domain. Very few clones were rescued from the first panning cycle after cosmix-plexing®, however, the second round showed a dramatic increase, demonstrating that the first round was sufficient to wash away the background phage (Figure III-29/B). After three rounds of panning, an enrichment factor of 570 was observed.

This strongly suggests that the recombined clones, selected after the first round of primary panning, were a good starting material for the cosmix-plexing®. Even if the enrichment obtained during the primary affinity selection could not be considered as significant, it has been now obviously demonstrated that the preselected population contained clones showing binding affinity for nephrocystin. This affinity was then further optimized during the cosmix-plexing® recombination, as proved by the output, and enrichment values of the secondary panning procedure (Figure III-29/A).

Randomly picked clones were sequenced, and provided clones which had not occurred during the primary panning selection (Figure IV-6). This suggests that the newly isolated variants have a selective advantage for nephrocystin, compared to the three clones previously described, this also being supported by phage ELISA analysis (Figure III-31). Clones isolated against nephrocystin, but also against Src, are able to interact with nephrocystin (Figure IV-6).

Half of the clones identified during the cosmix-plexing® panning, do not contain the proline rich cassette introduced in the CPLPPXP library, and are reduced to 8 amino acid long

peptides. It was possible to isolate such short peptides, due to the two steps protocol used to generate the CPLPPXP library. A primary bank was created with the insertion of a 8 amino acid encoding oligonucleotide, and then, the final CPLPPXP library was produced by the introduction of the proline rich cassette within the hypervariable oligonucleotide. When the cassette was ligated with the vector, 20% of vectors were shown to religate without cassette, forming back the parental vectors, containing the 8 amino acid peptide (Figure III-11).

Half of the nephrocystin selected clones after cosmix-plexing®, were selected from only 20% of the library, whereas for Src, none of the 8 residue variants was isolated, indicating that the shorter clones were providing a strong selective advantage for nephrocystin.

The analysis of the clones shows that nephrocystin ligands, contrary to Src ligands, do not necessarily require a **Pro Pro Ψ Pro** core motif, but prefer a **Pro Ψ Xxx Pro** motif. This latter motif is not present in the CPLPPXP library variants containing the proline rich cassette, explaining that the GST-nephrocystin target interacts preferentially with the 20% of the variants which did not contain the fixed proline motif.

Furthermore, since SH3 ligands required a motif of only 6 to 7 amino acids, the 8 amino acid variants displayed on the surface of 20% of the phage particle of the CPLPPXP library, are sufficient.

The second major variable between Src and nephrocystin ligands, involves the P_3 to P_5 motif. Src ligands were demonstrated to require at the positions P_3-P_5 the Arg Pro Leu motif. In clones where the proline was replaced by another amino acid, no strong interaction was possible with Src. However, nephrocystin ligands require a Arg Xxx Leu motif. A proline at P_4 is not disturbing, but is also not especially advantageous. The highest affinity nephrocystin ligand isolated (N5), according to phage ELISA test, exhibits a methionine at P_4, along with a Pro Val Tyr Pro core motif.

The only clone isolated on nephrocystin SH3 domain, but not able to interact with the target, was the N9 clone. N9 is the only reported clone (to our knowledge), isolated on SH3 domain, bracketed by two cysteines, and containing a proline core motif. As described in the section A of the discussion, panning on SH3 domains, with constrained libraries, failed to isolate SH3 ligands. The only reported peptide constrained by two cysteines was isolated against Src, through a “mirror image phage display” (Schumacher et al., 1996). However, this peptide is composed of non naturally occurring D-amino acids, and shows no similarity with characterized SH3 ligands (no proline rich motif). This suggests that disulfide closed loop

peptides would adopt a conformation different from the polyproline type II helix usually required for interactions with SH3 domains. The disulfide bridge was demonstrated to be crucial for interaction between this D-peptide and Src SH3 domain, indicating that the constraints imposed by both cysteines on D-peptide give a correct conformation to the peptide.

The N9 clone, which contains two cysteines, and the proline rich core motif, might not be able to form a conformation optimal for SH3 binding.

A clear consensus sequence was obtained from strongly binding nephrocystin ligands, however, a minimal consensus sequence showing the absolute requirements of nephrocystin ligands, can also be conceived, which takes into account only residues which are conserved in each high affinity ligand. The Src consensus sequence has very high homology with the nephrocystin consensus sequence, explaining the common cross-reaction of Src and nephrocystin ligands, for both SH3 domains. However, Src has more precise ligand preferences and is therefore more selective; more Src ligands bind to nephrocystin, than nephrocystin ligands to Src SH3 domain. The consensus sequences are presented here:

Consensus sequence from clones interacting with nephrocystin (Figure IV-6)

β Xxx **Arg (Pro) Leu Pro P/Ψ (Ψ) Pro P/V** Xxx **Pro** Xxx

Nephrocystin minimal consensus sequence (shown by all high affinity clones)

Arg Xxx **Leu Pro Ψ** Xxx **Pro**

Src consensus sequence

Arg Pro Leu Pro P/L Ψ Pro Ψ Ψ Pro S/Ψ

The ligands characterized against nephrocystin SH3 domain elevate the debate about biased libraries. Biased libraries are considered advantageous when working with targets of known ligand preferences, since they increase the chance of finding ligands with affinity of their displayed peptides for the target. However, the results obtained for nephrocystin demonstrate the potential risk in using such libraries.

We designed a biased library aimed to isolate strong ligands against SH3 domains, and it was chosen to define a PPXP proline core motif within the peptides of our CPLPPXP library. The library has been shown to be adequate in characterizing Src ligands, however, panning selection with nephrocystin revealed that a PXXP core motif would have been more appropriate for this target. To acquire ligands containing this required proline motif, the

nephrocystin had to interact with the 20% of the CPLPPXP library which show no fixed proline, leading to enrichment of variants exhibiting two prolines separated by two other amino acids. If the CPLPPXP library would have been designed by insertion of a unique sequence encoding the proline core motif within the hypervariable region (as done by several other groups), the bank would not have contained the optimal sequences, and the 8 residue PXXP ligands would not have been isolated. On this basis, it can be postulated that many potential ligands, against a large number of target proteins might have been lost, by a too systematic use of biased libraries containing a too highly defined motif.

	Sequences against Nephrocystin SH3 domain													E	B
	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}	P_{13}		
Strong binding clones															
N5			Arg	Met	Leu	Pro	Val	Tyr	Pro	Pro	*Asp*	*Pro*		+	+
S4	Lys	Gly	Arg	Pro	Leu	Pro	Pro	Val	Pro	Gly	Thr	Pro	Ser	+	+
N8			*Arg*	Pro	Leu	Pro	Pro	Thr	Pro	Pro	Ile	Pro	Gly	Ser	+	+
N10/S1	Gly	Asn	Arg	Pro	Leu	Pro	Pro	Ile	Pro	Ser	His	Pro	Phe		+	+
S2			Arg	Pro	Leu	Pro	Leu	Pro	Pro	Val	Pro	Trp	Val	Arg	+	+
N3*	His	Ser	Arg	Gln	Leu	Pro	Pro	Lys	Pro	Val	Pro	Ser	Leu		+	-
N4	*Arg*	Thr	Arg	Arg	Leu	Pro	Val	Leu	Pro	*Asp*	*Pro*				+	-
Consensus sequence															
	β	Xxx	Arg	(Pro)	Leu	Pro	P/Ψ	(Ψ)	Pro	P/V	Xxx	Pro	Xxx		
Weak binding clones															
N2*	Ser	Val	Arg	Leu	Leu	Pro	Pro	Val	Pro	Trp	His	Phe	Gly		+	-
N1/S3*			Arg	Ser	Leu	Pro	Leu	Pro	Pro	Val	Pro	Thr	Ser	Ser	+	-
N6			*Arg*	Thr	Pro	Pro	Thr	Glu	Pro	Gln	Thr				+	-
N7			Arg	Met	Leu	Pro	Val	Lys	Ile	Tyr					+	-
N11	Arg	Ala	Arg	Val	Val	Pro	Pro	Val	Pro	Arg	Leu	His	Thr		+	-
N9	Cys	Ser	Gln	His	Asp	Pro	Pro	Phe	Pro	Trp	Arg	Val	Cys		-	-

Figure IV-6: Sequences of single clones isolated against nephrocystin SH3 domain, and the deduced consensus sequence. The clones are classified according to their relative affinity on phage ELISA, for the target. Only the strongly positive clones in phage ELISA (Figure III-31), are taken into account for the conception of the consensus sequence. A residue is considered as representative for the consensus sequence, when it occurs in 50% or more of the sequences. Underlined amino acids are fixed residues within the CPLPPXP library. Ψ and β represent hydrophobic and basic amino acids, respectively. E and B stand for phage ELISA and BIAcore analysis. A simple + or – indicates whether the clones reacted positively or not, in ELISA or BIAcore tests. P_1 to P_{13} refer to the amino acid position within the sequence. The red asterisks indicate clones characterized during standard panning procedure. N clones were isolated against Nephrocystin, and S clones, against Src. Two clones (N10/S1 and N1/S3) were isolated against both nephrocystin and Src. For occurrence of each clone, or precise phage ELISA values, refer to Figures III-28, III-30 and III-31.

Ligand preferences for nephrocystin SH3 binding appear to be slightly different than that reported for SH3 domains. Until now, it has been thought that SH3 domains bind to proline rich ligands. Therefore, the use of biased libraries appeared to be an advisable approach to characterize SH3 ligands. However, it was recently reported that the SH3 domain of Eps8, a substrate of receptor and non receptor tyrosine kinases, displays novel binding preferences (Mongiovi et al., 1999). Eps8 SH3 domain interacts with ligands containing a PXXDY consensus sequence, with a K_D of 35 µM, similar to that found for SH3 domain/SH3 ligand interaction. Use of proline rich biased libraries with Eps8 would have lead to unsuccessful panning selection.

Biased libraries have to be carefully used. The ideal way to use a biased library, would be first, to isolate ligands against the target of interest, from a random peptide library. Then, in the eventuality that a consensus sequence can be defined, a biased library may be created, which contains the defined motif. This would minimize the risk of eliminating the high affinity ligands for a particular target, within a biased library. Biased libraries should be regarded as secondary libraries.

Cosmix-plexing® libraries are optimal for such strategies. Used as primary random libraries to allow determination of a consensus sequence, they can then be converted into biased libraries, since they are designed for facile insertion of a project specific cassette within the hypervariable domain.

1-2-2 BIAcore analysis of nephrocystin ligands

The five clones which showed the highest relative binding on phage ELISA are also reacting positively on BIAcore real time interaction measurements (N5, N8, N10/S1, S2, S4; Figure III-47/B). Their K_D were determined by steady state affinity data evaluation, with BIAcore measurements, and ranged between 0,42 and 1,05 nM (Figure III-51).

When compared with usual K_D values of SH3 ligands (Table IV-4), the nephrocystin ligands are the highest affinity ligands ever characterized. Nephrocystin ligands appear to interact with nephrocystin with an affinity 20 fold higher than for the CPLPPXP S1 clone/Src interaction, and approximately 500 fold higher than the best SH3 ligands ever characterized before this work.

The S1 clone binds nephrocystin with different kinetics compared to other nephrocystin ligands, also different from the way S1 binds to Src (Figure III-47). S1 shows a extremely rapid association with nephrocystin, compared to other clones, but the dissociation is also

very fast, while other clones form a more stable complex with nephrocystin. However, their K_D values are very similar (0,52 nM for S1 and 0,42 for S4), and do not reflect the different binding features. S1 and S4 show similar dissociation constants, however, their association and dissociation rate constants are different. It might therefore be of general interest to evaluate the k_a and k_d along with the K_D. Kinetic information about biospecific interactions can complement affinity measurements. S1 and S4 interactions with nephrocystin, present different interests for *in vitro* analysis. Their distinct behaviors in interacting with nephrocystin SH3 domain might compete differently other ligands.

1-2-3 Swissprot comparison of nephrocystin ligands

A computer-assisted sequence search with the Swissprot database revealed homology between several nephrocystin ligands and naturally existing proteins.

The Table IV-6 presents the diverse homologies of the nephrocystin ligands.

N1 ligand Rat CAP1	RSLPLPPVPTSSA PGPPPPPVPTSSG	89% identity in 9 aa overlap CAP1: aa 237-245
N1 ligand Human WASP	RSLPLPPVPT SSA RSGPLPPVPLGIA	70% identity in 13 aa overlap WASP: aa 338-350
N3 ligand Mouse WASP Human WASP	HSRQLPPKPVPSL ERRQLPPP PAPIN DRRQLPPPPTPAN	77,7% identity in 9 aa overlap WASP: aa 159-167
N8 ligand Mouse WASP Human WASP	RPLPPTPP IPGSGS PP PPPPPPCPGSGP PP PPPPPSSGNGP	75% identity in 12 aa overlap WASP: aa 416-427
N8 ligand Human RHG1	RP LPPTPP I PGSGS KPMPPRPPLPNQQF	60% identity in 10 aa overlap RHG1: aa 229-238
N4 ligand Human Sck	TRRLPV LP TRRAPVAP	75% identity in 8 aa overlap Sck: aa 357-364
N6 ligand HIV1 Nef	RTPPTEPQT RTPPTETGV	100% identity in 6 aa overlap Nef: aa 22-27

Table IV-6: Swissprot database analysis of nephrocystin ligands. Regions of identity are shown in red, and of homology in blue.

N1 (also called S3) shows 70% identity in a stretch of 13 amino acids with the human WASP (Wiskott-Aldrich Syndrome Protein), and 89% identity in a stretch of 9 residues, with the rat CAP1 protein (Adenylyl cyclase associated protein 1). N1 homologies are analyzed under section B-1-5 of the discussion, as homologies of the S3 clone.

N3 and N8 also show homologies to WASP: 77,7% identity in a stretch of 9 amino acids, and 75% identity in a 12 residue stretch, respectively. In addition, N8 has 60% identity in a 10 amino acid stretch, with the human Rho-GTPase activating protein 1 (RHG1).

N4 shows a 75% identity with the human Sck, in a 8 residue stretch.

And finally, N6 shows a 100% identity with the HIV-1 Nef protein, in a 6 amino acid stretch.

The identity between the N1 (also called S3) clones and the CAP1 protein was previously discussed under section A from discussion, and therefore, is not repeated here.

1-2-3-1 WASP homologies

Homologies of N1, N3 and N8 for WASP appear very curious, since the three ligands have identity with various sections of the WASP protein.

Defective WASP is associated with the Wiskott-Aldrich syndrome, leading to death usually before the age of 10, and characterized by platelet and lymphocyte cell abnormalities, most likely caused by defects in actin cytoskeleton organization (Ochs et al., 1980).

WASP has been shown to interact with many SH3 domains, Itk, Src, Lck, Abl, Nck3 and Grb2 (Bunnell et al., 1996; Rivero-Lezcano et al., 1994; She et al., 1997). Bunnell et al. (1996) also demonstrated that WASP contains distinct binding sites for its interactions with Itk, Src, Lck and Abl SH3 domains. The N1 (S3) homology with WASP was demonstrated to concern a region a WASP involved in interaction with C terminal SH3 domain of Nck. However, the WASP binding sites for the other SH3 domains were not clearly defined, and further homology analysis is not possible.

1-2-3-2 rho-GTPase activating protein 1 homologies

In addition to its similarity with a region of the WASP protein, the N8 ligand shows 60% identity in a stretch of 10 residues, with the human RHG1. Like the nephrocystin N8 ligand, the 3BP-1 protein, which was the first isolated SH3 ligand, showing 71% identity over a 17 amino acid region with the RHG1 (former described as GAP-rho) which participate in cytoskeletal organization (Cicchetti et al., 1992). However, N8 and 3BP-1 present homologies

with a different region of the RHG1.

The function of the RHG1, as its full name indicates, is to activate the GTPases of the rho, Rac, and cdc42 proteins, and to convert them to the putatively inactive GDP-bound state. Cdc42 appears to be its preferred substrate. RHG1 contains a functional SH3 binding domain (Barfod et al., 1993), however, the N8 ligand homology for RHG1 does not concern the SH3 binding site (aa 251 to 262), but a region short before (aa 229 to 238).

The Rho family member cdc42 interact with multiple proteins belonging to the family of the serine/threonine kinases. Interaction of cdc42 with Ser/Thr kinases, activates their kinase domain. Ser/Thr kinases can then regulate the MAP kinase signaling cascades, leading to actin cytoskeletal rearrangements.

The activated cdc42 protein has been shown to interact with WASP (Symons et al., 1996).

Homologies for nephrocystin ligands with WASP or RHG1, even if not always clear, seem to indicate a link between nephrocystin SH3 domain, and the actin cytoskeleton.

1-2-3-3 Sck homologies

The nephrocystin N4 ligand show 75% identity on a region of 8 residues with the human Sck protein. Sck contains a SH2 domain (Kavanaugh et al., 1994). The homology between N4 and Sck suggests that Sck could be recruited by Src SH3 domain, or the contrary.

1-2-3-4 HIV-1 Nef homologies

The nephrocystin N6 ligand also shows complete identity (100%) with a 6 amino acid region of the Human Immunodeficiency Virus type 1 (HIV-1) Nef protein. Nef is required for enhancing viral infectivity, however, its *in vivo* role in pathogenicity remains unclear (Lee et al., 1996). Nef acts at an early step of the viral replication cycle. The Nef protein has been shown to interact with a subset of Src family kinase SH3 domains, with a very high affinity (K_D of 0,25µM for Hck SH3; Saksela et al., 1995). The physiological relevance of the interaction of Nef with SH3 domains is unknown. Nef has also been shown to bind a serine/threonine kinase, through a RR motif in Nef, and it appears that other cellular factor(s) are involved in the efficiency of the viral entry (Tokunaga et al., 1999). However, the identity is unlikely to be significant, since Nef binds SH3 domain at a different region than the one showing homology with the N6 ligand.

1-3 Nck SH3 domains

Difficulties appeared when panning selections were performed on N terminal and central Nck SH3 domains (Nck1 and Nck2). After standard affinity selections, no enrichment was observed. However, sequencing of picked clones led to the determination of a short consensus sequence, **Asp Gly Tyr**, conceived from only two clones which occurred several times, for both Nck1 and Nck2.

NP1	Ser Thr Leu **Cys Asp Gly Tyr Cys**
NP2	Asn Pro Met Gly **Asp Gly Tyr** Phe

Curiously, these two clones do not present a proline motif (so, both clones were called Non Proline 1 and 2 clones: NP1 and NP2), and belong to the 20% of the CPLPPXP library which does not contain the proline rich cassette.

A cosmix-plexing® recombination, followed by screening of the secondary library were performed, and as far as Nck1 was concerned, gave the same results as for the primary panning, only NP1 and NP2 were found.

For Nck2, even though no significant enrichment was detected, randomly selected clones allowed the characterization of clones corresponding to expected SH3 ligands, containing a proline rich core motif. The clones occurred only once each, indicating that none of them showed a dominant affinity for Nck2 over the other variants. However, surprisingly, the alignment of the clones show some homologies. The variants were divided into three groups, according to their similarities. All three sets present an over-representation of aromatic residues, especially tryptophan and tyrosine, directly after the proline core motif (P_{10}), but differ in other characteristics (Figure IV-7).

The first group of clones were grouped together by virtue of the proline residue at P_{12}. Other similarities were obvious; Phe at P_8, Cys or Thr at P_3, or hydrophobic residues at P_4.

Clones of the second group were selected because of their arginine at P_5, but other homologies were observed, such as Ser at P_1 and Ser or Cys at P_{13}, and Leu at P_{11}.

The third group contained the clones which did not belong to the previous two classifications. However, even though no position was defined as a particular residue, except the aromatics at P_{10}, a very high content of hydrophobic residues was observed.

However, after a further cycle of panning, NP1 and NP2 largely dominate other clones.

Sequences against Nck2 SH3 domain, from cosmix plexing library

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}	P_{13}
First group													
	Ser	Leu	Gly	**Ile**	Trp	**Pro**	**Pro**	Phe	**Pro**	Trp	Ile	**Pro**	**Ile**
	Ala	Ser	Cys	**Phe**	Gln	**Pro**	**Pro**	Phe	**Pro**	Trp	Thr	**Pro**	**Leu**
	Arg	His	Cys	**Val**	Gly	**Pro**	**Pro**	Asp	**Pro**	Tyr	Val	**Pro**	**Met**
	Thr	Cys	**Thr**	**Ala**	Leu	**Pro**	**Pro**	Phe	**Pro**	Trp	Ser	**Pro**	Arg
	Leu	Val	**Thr**	Cys	Val	**Pro**	**Pro**	Phe	**Pro**	Trp	Pro	**Pro**	**Pro**
Consensus sequence													
	Xxx	Xxx	**C/T**	**Ψ**	Xxx	**Pro**	**Pro**	Phe	**Pro**	W/Y	Xxx	**Pro**	**Ψ**
Second group													
	Pro	Arg	Asn	Val	**Arg**	**Pro**	**Pro**	Ala	**Pro**	Trp	His	Gly	**Ser**
	Ser	Thr	Gly	**Arg**	**Arg**	**Pro**	**Pro**	His	**Pro**	Phe	**Leu**	Val	**Ser**
	Lys	Phe	Met	Asn	**Arg**	**Pro**	**Pro**	Gly	**Pro**	Trp	**Leu**	Ala	Cys
	Ala	Ser	Ser	Thr	**Arg**	**Pro**	**Pro**	Cys	**Pro**	Tyr	**Leu**	Asp	Ile
	Arg	Pro	Arg	**His**	**Arg**	**Pro**	**Pro**	Val	**Pro**	Val	**Val**	Lys	Pro
	Ser	Val	Phe	Leu	**Arg**	**Pro**	**Pro**	Cys	**Pro**	Ala	**Leu**	Gly	Asn
	Ser	Ser	Tyr	**Arg**	**Arg**	**Pro**	**Pro**	Leu	**Pro**	Ser	His	Arg	Cys
Consensus sequence													
	(Ser)	Xxx	Xxx	**(β)**	**Arg**	**Pro**	**Pro**	Xxx	**Pro**	**φ**	**Leu**	Xxx	**S/C**
Third group													
	Met	Phe	Gln	**Phe**	Gly	**Pro**	**Pro**	Cys	**Pro**	Tyr	**Leu**	Asp	**Ile**
	Val	Gly	**Leu**	His	Trp	**Pro**	**Pro**	**Val**	**Pro**	Trp	**Val**	Gln	Arg
	Leu	Lys	**Leu**	**Val**	Asn	**Pro**	**Pro**	Gln	**Pro**	Tyr	**Phe**	**Ile**	**Leu**
	Asp	Ser	**Leu**	**Ala**	Glu	**Pro**	**Pro**	**Phe**	**Pro**	Tyr	**Met**	**Ile**	**Val**
	Gly	Glu	Tyr	**Val**	Trp	**Pro**	**Pro**	**Val**	**Pro**	Trp	Ser	**Met**	**Met**
Consensus sequence													
	Ψ	Xxx	**Leu**	**Ψ**	Xxx	**Pro**	**Pro**	**Ψ**	**Pro**	W/Y	**Ψ**	**Ψ**	**Ψ**

Figure IV-7: Sequences of single clones, after two rounds of cosmix-plexing® panning against Nck2 SH3 domain, and the deduced consensus sequences. The clones occurred only once each. A residue is considered as representative for the consensus sequence, when it occurs in 50% or more of the sequences. Underlined amino acids are fixed residues within the CPLPPXP library. Ψ, β, and φ represent hydrophobic, basic and aromatic amino acids, respectively. P_1 to P_{13} refer to the amino acid position within the sequence.

It was reported that strong negative selection commonly occurs against odd number of cysteines residues (Kay et al., 1993). It was hypothesized that phage that express unpaired or lone cysteines may be detrimental to phage propagation, since the cysteines are likely to form disulfide bridges with cysteines of other displayed peptides, or with cysteines of the phage.

The pIII protein contains 8 cysteines, which are all involved in disulfide bridges (Kremser and Rashed, 1994). Displayed peptides containing unpaired cysteines, might connect with one of the cysteines of the pIII, and therefore, disturb the conformation of the complete particle.

However, the clones selected against Nck2 SH3 domain show an extremely high frequency of lone cysteines. From the first two groups of variants, 66% of the clones contained a single cysteines. None of the clones possess a pair of cysteines. Within the first group, the cysteines are located N terminally from the proline rich motif, while in the second group, the cysteines are localized C terminally. It has to be pointed out that stable phage displaying peptides containing a single cysteine have already been reported (Armstrong et al., 1996). Presumably, these lone cysteines were buried in an inaccessible region of the peptides which fold up on themselves.

The reason of the presence of single cysteines within Nck2 SH3 domain is unclear. It might be possible that these cysteines form a disulfide bridge with a cysteine of the Nck2 protein.

It was concluded that cosmix-plexing® recombination allowed the isolation of proline rich clones which could be grouped in three consensus sequences. However, these clones were not quantitatively enriched, and failed to react with the Nck2 SH3 domain on phage ELISA, indicating that the clones if really interacting with Nck2, were binding too weakly to give a signal. The only clones giving a signal in phage ELISA on Nck1 and Nck2 SH3 domains, were the non prolines NP1 and NP2 clones.

One consideration would be that the incapacity of isolating strong proline rich ligands against Nck1 and Nck2 would be explained by the fact that NP1 and NP2 competed effectively with other clones. However, this is unlikely, since NP1 and NP2 react moderately, on phage ELISA, with Nck1, and very weakly with Nck2 SH3 domains.

In the light of the results obtained for nephrocystin, it might be possible that failure to isolate high affinity clones for the various Nck SH3 domains is due to a non appropriate use of biased libraries, as discussed in the previous section. It is possible that the PPXP core motif imposed to the CPLPPXP library, is not advantageous for Nck SH3 domains. Furthermore, reported Nck ligands seem to be longer than for other SH3 domains (10 to 12 amino acid long regions), reducing the chance that Nck SH3 domains interact with some of the 8 amino acid random peptides, constituting 20% of the CPLPPXP library. However, these Nck ligands are

natural binders, and peptide ligands, isolated from phage display might be different, or present a minimal core motif necessary for interaction with Nck, which would consists only of a portion of known ligands.

Isolation of ligands for one of the Nck SH3 domains using phage display has not been reported. Spark et al. (1996) described the expression of GST constructs, where the GST was independently fused to the Nck1, Nck2 and Nck3 SH3 domains, however, no results on Nck have been presented, suggesting a failure in the panning procedure. Sparks et al. were also using a biased library (X_6PXXPX_6).

Nck has been shown to interact with a large variety of proteins. The three SH3 domains of Nck (Nck1, Nck2, Nck3) can interact selectively with target proteins. The Nck1 domain mediates Nck association with the Nck-Associated Protein 1 (Nap1) (Kitamura et al., 1996).

Nck2 domain links Nck with many ligands, including p21-Associated Kinase (PAK) (Galisteo et al., 1996; Lu et al., 1997), Sos, a guanidine nucleotide exchange factor for Ras (Hu et al., 1995), the serine/threonine kinase PRK2/NAK (Quilliams et al., 1996; Chou and Hanafusa, 1995), Nck, Ash-, and phospholipase Cγ-binding protein 4 (Matuoka et al., 1997), and WIP (WASP Interacting Protein; Anton et al., 1998).

The Nck3 domain mediates Nck interaction with the Wiskott-Aldrich syndrome protein (WASP) (Rivero-Lezcano, et al., 1995), and the γ isoform of the serine/threonine kinase casein kinase I (CKI-γ2) (Lussier et al., 1997).

Other proteins that have been shown to interact with Nck, but for which the specific SH3 domain that mediates binding has not been defined, include c-Cbl (Rivero-Lezcano et al., 1994), focal adhesion kinase (Choudhury et al., 1996), pp105, a lymphocyte-type CRK-associated substrate that bind to FAK and Crk (Minegishi et al., 1996), and Nck Interacting Kinase (NIK; Su et al., 1997).

Despite the number of possible Nck interactions, no phage display isolated ligands were reported. In this work, as well as in Sparks (Sparks et al., 1996), interactions with individual SH3 domains of Nck were investigated. However, presence of three distinct SH3 domains suggests that Nck may be capable of simultaneously associate with several ligands, or that binding partners require more than one SH3 domain.

Dynamin selectively interacts with one of the C terminal Nck SH3 domain (Wunderlich et al., 1999a). Similarly the PRK2 binds to the second SH3 domain of Nck (Quilliams et al., 1996). In both cases, the proteins bind the individual Nck SH3 domains, and the full length

Nck with similar affinity.

However, Sos (Wunderlich et al., 1999a), WASP (Finan et al, 1996) and PAK1 (Bokoch et al., 1996) were demonstrated to bind specifically with the third (Sos, WASP) or second (PAK1) Nck SH3 domains, but, for the three proteins, combined SH3 domains contributed to a high affinity binding.

Cbl binds to constructs containing the first and the second Nck SH3 domains, however, it do not interact with any of the individual SH3 domains. Proteins like Cbl, Bcr/Abl, p155, and NIK bind *in vitro*, only to combined SH3 domains (Wunderlich et al., 1999b).

Since it is proved that Nck SH3 domain cooperation enhances the binding of proteins (Wunderlich et al., 1999b), work with full length Nck would have enhanced the chances of isolating high affinity ligands.

Furthermore, it seems that Nck ligands contain several proline rich domains (e.g. WASP, WIP, Sos, NIK) and therefore it could be that some Nck ligands interact simultaneously with several Nck domains. If multiple interactions are required, short peptides are not relevant to reproduce such interactions, and would fail to bind to Nck with high affinity.

2-EVH1 domains

In 1997, it was demonstrated for the first time that the EVH1 domains interact with proline rich ligands (Niebuhr et al., 1997). Since then, an EVH1 domain-containing protein family has been characterized, and several new binding targets were discovered, allowing specification of their ligand preferences. However, to our knowledge, this is the first time that phage display was used to isolate ligands against various EVH1 domains.

2-1 Vesl/Homer EVH1 domain

Vesl (also called Homer) belongs to the family of EVH1 domain containing proteins, but structurally, and functionally, appears to be different (Figure IV-8). While EVH1 domain binding mediates proper localization of Ena, VASP, Mena, Evl, and WASP, which then recruit components of the actin assembly machinery (Prehoda et al., 1999), Vesl functions in synaptic plasticity (Brakeman et al., 1997; Kato et al., 1997).

The variations within the consensus sequences of EVH1 ligands, reflect the functional differences between Vesl and the other members of the EVH1 family. EVH1 domains involved in actin dynamics recognize ligands with a **Phe Pro Pro Pro Pro Pro** consensus sequence, while Vesl interacts with ligands containing the **Pro Pro Xxx Pro Phe Arg Asp**

consensus sequence, which look quite similar to the SH3 class II ligands (Pro Pro Xxx Pro Xxx Arg) (Figure III-33).

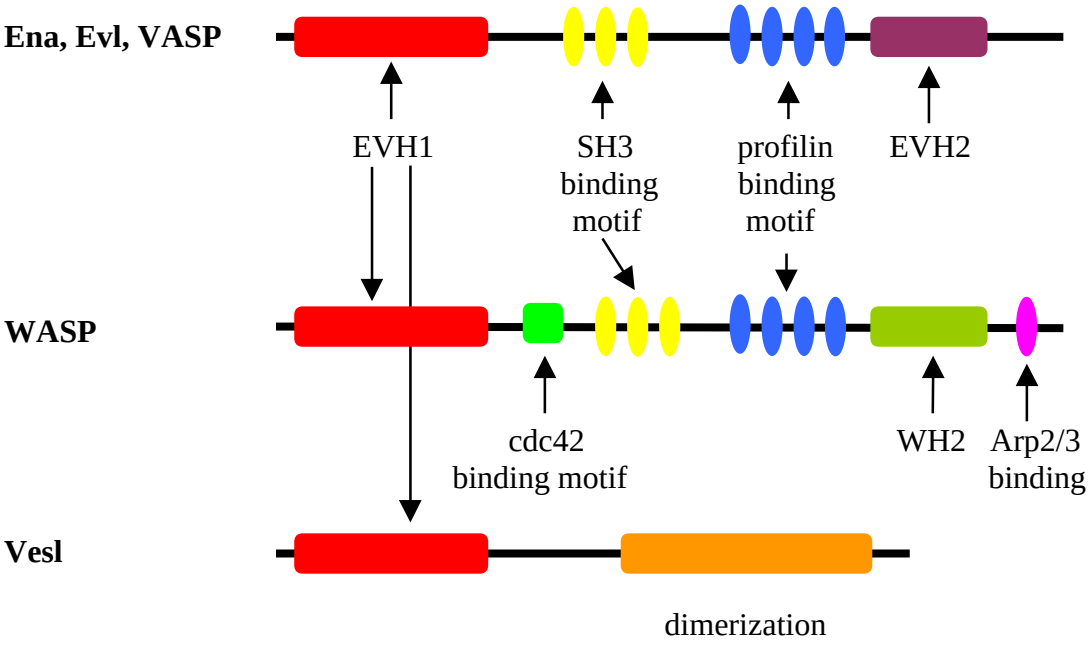


Figure IV-8: Illustration of the structure of the EVH1 domain containing proteins. Ena, Evl, VASP, and WASP are involved in regulation of the actin cytoskeleton and contain motifs that can interact with various signaling proteins, and actin binding proteins. The Vesl plays a role in synaptic plasticity (Prehoda et al., 1999).

Panning selection was performed on Vesl, according to the protocol optimized for Src SH3 domain. Standard panning led to high enrichment (600) after four cycles, and to the obtention of a consensus sequence: **Arg Ψ** (**φ**/**Ψ**) Xxx **Pro Pro** (**φ**/**Ψ**) **Pro W/Y** Xxx **D/F**

The cosmix-plexing® recombination, followed by affinity selection, led to an enrichment factor of 1000 after three rounds of panning. Clones sequenced from the first round showed predominance of the V1 clone, which was already the most frequent variant after standard panning, but in addition, several new clones were isolated, which clarify the consensus sequence: **Ser Arg Ψ Y/F (α) Pro Pro F/P Pro W/Y** Xxx **Ψ Ψ**

However, the alignment of clones which interact strongly with Vesl, on phage ELISA (Figure III-38), led to the consensus sequence which reflects the strongest ligand preferences of Vesl: **Arg Ψ Y/F (Asp) Pro Pro F/Ψ Pro W/Y** Xxx **Asp Ψ**

The clones isolated against Vesl are presented in Figure IV-9, according to their relative binding for Vesl. The essential positions for high affinity binding with Vesl EVH1 domain are most of all, the aromatic residue (tryptophan or tyrosine) at P_{10}, along with Arg at P_2 and an

aromatic residue (tyrosine, or eventually phenylalanine) at P_4. The aspartic acid at position P_{12} appears to be advantageous, since the three best clones have it, however, it is not crucial for the interaction. It may favor the binding between Vesl and its ligands, through the formation of a salt bridge, which would stabilize the interaction.

	Phage ELISA on Vesl EVH1 domain (positive clones)												OD_{492}
	P_1	P_2	P_3	P_4	P_5	P_6 P_7		P_8	P_9	P_{10}	P_{11}	P_{12}	P_{13}	
Strong binding clones														
V1	Asp	**Arg**	His	**Tyr**	Arg	**Pro**	**Pro**	**Phe**	**Pro**	**Trp**	Ala	**Asp**	Gly	0,87
V2	Asn	**Arg**	**Leu**	**Tyr**	Pro	**Pro**	**Pro**	**Trp**	**Pro**	**Tyr**	Ser	**Asp**	**Pro**	0,76
V6	Tyr	Leu	**Ile**	**Leu**	Ser	**Pro**	**Pro**	**Ala**	**Pro**	**Trp**	Arg	**Asp**	Arg	0,67
V3	Ser	**Arg**	Ser	**Val**	Tyr	**Pro**	**Pro**	**Pro**	**Pro**	**Tyr**	Pro	Phe	**Ala**	0,55
V8	His	Met	**Leu**	**Phe**	Asp	**Pro**	**Pro**	**Phe**	**Pro**	**Tyr**	Ser	Asn	Glu	0,52
V9	Ser	**His**	**Ile**	**Phe**	Asp	**Pro**	**Pro**	**Phe**	**Pro**	**Tyr**	Gly	Pro	**Met**	0,48
Consensus sequence														
	Xxx	**Arg**	**Ψ**	**Y/F**	**(Asp)**	**Pro**	**Pro**	**F/Ψ**	**Pro**	**W/Y**	Xxx	**Asp**	**Ψ**	
Weak binding clones														
V4	Asp	**Arg**	**Ile**	**Tyr**	Leu	**Pro**	**Pro**	**Val**	**Pro**	**Trp**	Ala	Asn	Ser	0,21
V11	Ser	**Arg**	**Val**	**Tyr**	Pro	**Pro**	**Pro**	**Phe**	**Pro**	**Trp**	Arg	Ala	**Val**	0,17
V7	Cys	Ser	Arg	Ser	Arg	**Pro**	**Pro**	**Val**	**Pro**	Leu	Gly	Pro	**Phe**	0,05
V5	Arg	Thr	Gly	**Val**	Trp	**Pro**	**Pro**	**Pro**	**Pro**	His	Asp	Phe	Arg	0,03
V10	Ser	Leu	**Leu**	**Tyr**	Glu	**Pro**	**Pro**	**Pro**	**Pro**	**Trp**	Asn	Ser	**Pro**	0

Figure IV-9: Relative affinity, according to a phage ELISA test, of the clones isolated against Vesl EVH1 domain. The clones are classified according to their phage ELISA signals. None of the clones react with GST alone (not shown). A new consensus sequence is conceived from the strongest clones. Underlined residues are fixed within the CPLPPXP library. Ψ represents hydrophobic residues.

Comparison with the consensus sequence deduced from alignment of clones isolated from phage display technology, with the consensus sequence obtained from alignment of Vesl natural ligands shows homology.

Phage display consensus sequence										
	Arg	**Ψ**	**Y/F**	**(Asp)**	**Pro**	**Pro**	**F/Ψ**	**Pro**	**W/Y**	Xxx	**Asp**	**Ψ**
Natural ligand consensus sequence										
	Xxx	**Pro**	**Pro**	Xxx	**Pro**	**Phe**	**Arg**	**Asp**		
Similarities between both										
	Pro	**Pro**	Xxx	**Pro**	**Arom**	Xxx	**Asp**			

The proline core motif, the aromatic position directly following, and the Asp three residues after the proline motif, are conserved between natural ligands and the clones isolated with phage display. However, natural ligands preferred a phenylalanine as aromatic amino acid, while the peptide ligands exhibit a tyrosine or a tryptophan. The only position defined for the natural ligands, but not for the peptides, is an Arg, between the aromatic residue, and the aspartic acid, on the C terminal part of the proline motif.

The consensus sequence deduced from the phage display ligands provides more information about the region N terminal to the proline core motif.

It is interesting to notice that the Vesl motif possesses similarities to the SH3 motif, especially, according to our results, to the ligand preferences of Nck2, which also possess an important aromatic residue directly after the core motif. The observation that Sos which is known as an SH3 ligand, can bind Vesl, using the same sequence for SH3 and Vesl EVH1 domains, confirms the link between SH3 domains and Vesl EVH1 domain ligand preferences (Prehoda et al., 1999).

As with studies of SH3 interactions, the Vesl ligand peptides should provide a valuable tool to assess the functions of Vesl.

A Swissprot analysis of the various clones obtained after affinity selection showed 80% identity between the V3 clone and the human MGR1 (metabotropic glutamate receptor 1 precursor), a receptor for glutamate, belonging to the mGluR5 family, a G protein coupled receptor.

V3 ligand	S**RS***V***YPPPP**Y**P**FA	80% identity in 10 aa overlap
Human MGR1	L**RS***L***YPPPP**P**P**QH	MGR1: aa 1063-1072

Group1 metabotropic glutamate receptors (mGluR) represent a family of seven membrane-spanning proteins, coupled to G proteins and which activate phospholipase C. Members of the family include mGluR5. Activation of these receptors activate protein kinase C, and inositol triphosphate (IP3) production, which induce the IP3 receptors to release intracellular calcium. Metabotropic signaling has been implicated in several forms of activity dependent synaptic plasticity, but its role remains controversial (Tu et al., 1998).

However, the mGluR5 binds Vesl, through a TPPSPFR region entirely conserved in the MGR1, on the C terminal domain, which is not the one concerned by the homology with the V3 clone.

2-2 Mena, Evl and VASP EVH1 domains

Standard panning selection on Mena, Evl and VASP EVH1 domains led to the isolation of clones, which were not quantitatively enriched, but which share homologies. However, the characterized variants did not react against the EVH1 domains, on phage ELISA. A large number of unproductive clones were found (40 to 70% of the sequenced clones, according to the proteins). In addition, the non proline clones NP1 and NP2 were isolated several times.

However, consensus sequences have been conceived for all three proteins, and are shown, along with the consensus sequence of natural ligands interacting with the EVH1 containing proteins, involved in actin assembly:

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}	P_{13}
Evl	**φ**	Xxx	**Ψ**	**Ψ**	**Ψ**	**Pro**	**Pro**	**Ψ**	**Pro**	**W/Y**	Xxx	Xxx	Xxx
VASP	**Ψ**	Xxx	**Ψ**	Xxx	**Ψ/φ**	**Pro**	**Pro**	**Phe**	**Pro**	**Tyr**	Xxx	**Y/F**	Xxx
Mena	Xxx	Xxx	Xxx	**Ψ**	**α/L**	**Pro**	**Pro**	**Ψ**	**Pro**	Met	**Y/L**	**Ψ**	**Ψ/Y**
Natural EVH1 family ligands													
	Xxx	*Xxx*	***Ψ***	***α***	***Phe***	***Pro***	***Pro***	***Pro***	***Pro***	***Pro***	***Ψ***	***α***	***Glu/Ψ***

The EVH1 ligand preferences observed with our clones, do not correspond to expectations, regarding the consensus sequence deduced from EVH1 domain natural ligands.

The main feature of EVH1 natural ligands, is the phenylalanine position directly before the proline core motif, at P_5. Individual mutations showed that the phenylalanine can be replaced by tyrosine or tryptophan, which even improved the binding between the natural ligand and the EVH1 domain. Every other amino acid at this position disrupted the interaction (Niebuhr et al., 1997). This aromatic position has been shown to be absolutely essential for the interaction. According to the peptide scan performed by Niebuhr et al. (1997), the proline replacements were quite well tolerated, at least if the prolines were mutated one at a time. However, prolines at positions P_7 and P_8 appeared to be more sensitive than at P_6 and P_9. Substitutions at P_{10} are do not alter the binding.

However, according to the consensus sequences obtained after primary panning, from the phage display clones, EVH1 characteristics appear as a mirror image of that observed with the natural ligands. Indeed, the most defined position, for the three proteins, is P_{10} (which was silent in natural ligands). Furthermore, this position is determined, for Evl and VASP, as an aromatic residue (tyrosine or tryptophan for Evl, tyrosine for VASP), while aromatic residues were found at P_5 in natural ligands. This crucial P_5 aromatic position of the natural ligands, appears mainly as hydrophobic positions in our clones, even if VASP seems to prefer

aromatic hydrophobic residues (tryptophan or phenylalanine).

This might suggest, as proposed by Tu et al. (1998), Mahoney et al (1999), and Fedorov et al. (1999), that the degeneracy observed for SH3 domains, and profilin might also be valid for EVH1 domains, and that they might be able to bind their proline rich ligands in two orientations. Fedorov et al. (1999) even suggested that the polarity would then be determined by the phenylalanine (at P_5 within the natural ligands, and at P_{10}, like in our clones, for the second class of ligands), which would interact with a number a methylene groups contributed by four conserved hydrophilic residues. However, this has not yet been proven, and no ligands containing the aromatic residue after the proline motif have been characterized previously.

A cosmix-plexing® recombination was performed, but screening of the secondary library led exclusively to the isolation of the non-proline clones, NP1 and NP2. These NP1 and NP2 variants were shown to specifically interact *in vitro* with the various EVH1 domain containing proteins. However, their binding shows target specificity. Vesl and Mena react strongly with the non proline clones; NP1 and NP2-VASP interactions are quite weak, whereas ELISA signals between the non proline variants and Evl EVH1 domain are almost non-existent, i.e. OD values are nearly at the background level (Figure III-40). Up to now, parallels between SH3 and EVH1 domains were limited to the observation that they both prefer proline-rich ligands adopting a polyproline type II helix. It was also suggested that both SH3 and EVH1 domains were able to interact with their ligands in two orientations (shown for SH3 domains, but not for EVH1 domains). However, the non proline clones provide evidence of a stronger link, since it is the first time that a common ligand between both modules is characterized (Niebuhr et al. (1997) showed that usual SH3 ligands were not binding EVH1 domains *in vitro*).

Since the consensus sequences obtained for Evl and VASP show homologies with those for Vesl, especially the aromatic residue at position P_{10}, it was decided to screen the cosmix-plexed® library optimized for Vesl, against the other EVH1 domains.

The affinity selection against Mena was unsuccessful, as expected since the Mena consensus sequence was not resembling that of Vesl (Mena binding does not require aromatic residue at P_5 as do Vesl, or at P_{10} like Evl and VASP, but rather at P_{11} or P_{13}).

However, this system allowed characterization of new clones against Evl, and VASP, which were in some cases quantitatively enriched. This permitted the conception of consensus sequences.

Evl clones could be classified into two groups, which allowed consensus sequences to be made. However, these consensus sequences are very different from those obtained from the original CPLPPXP library. They also show no similarity with the natural ligand consensus sequence.

Evl, first group	Xxx	Xxx	Xxx	**Ile**	Xxx	**Pro**	**Pro**	**Cys**	**Pro**	**Ser**	Xxx	**φ**	Xxx
Evl, second group	Xxx	Xxx	**Ψ**	**Ψ**	Xxx	**Pro**	**Pro**	**Phe**	**Pro**	**Asp**	**Ψ**	**Ψ**	Xxx

However, every clone identified with the Vesl optimized library, against Evl, possesses an acidic residue, at P_3, P_5, P_{10} (preferred) or P_{11} (Figure III-41), which is a particular characteristic of natural EVH1 ligands. However, none of the clones (even the dominant one isolate 7 fold) was able to interact *in vitro* with Evl, according to phage ELISA.

Clones identified against VASP also yielded to two consensus sequences, which showed homologies with the consensus sequence conceived after panning selection using the original CPLPPXP library.

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}	P_{13}
CPLPPXP library	**Ψ**	Xxx	**Ψ**	Xxx	**Ψ/φ**	**Pro**	**Pro**	**Phe**	**Pro**	**Tyr**	Xxx	**Y/F**	Xxx
VASP first group	**Asp**	**Arg**	**Leu**	**Tyr**	**α**	**Pro**	**Pro**	**F/Ψ**	**Pro**	**W/Y**	Xxx	Xxx	**Ψ**
VASP second group	Xxx	**Leu**	**β**	**Y/V**	**Arg**	**Pro**	**Pro**	Xxx	**Pro**	**Trp**	**Ψ**	Xxx	Xxx

All three consensus sequences present an aromatic residue at P_{10}. Otherwise, the new consensus sequences are longer, and more defined than the one obtained on standard panning, indicating that the screening performed with a library optimized for another target, but which showed strong homologies with VASP, was appropriate.

The Vesl optimized library consensus sequences show for both groups a high preference for tyrosine at P_4. This position might be a link between our clones, and the natural ligands, which all contain a phenylalanine directly before the core motif, which can be replaced by other aromatic residues (Niebuhr et al., 1997).

For both newly obtained consensus sequences, the region N terminal from the proline motif is more defined than the C terminal part.

More charged residues occurred within the sequences. 8 out of the 9 isolated clones (Figure III-42), possessed one or two acidic residues (especially within the first group). This is important, since EVH1 ligands are assumed to stabilize their interaction with the domains, through salt bridge contacts between their acidic residues, and conserved basic amino acids of the EVH1 domains. A high frequency of basic residues are also observed, especially for the variants of the second group. However, their function is unclear.

Nevertheless, in spite of the homologies between the clones, and with previous consensus sequences, the variants were unable to interact with VASP *in vitro*, on phage ELISA.

3- Non proline ligands

3-1 NP1 and NP2 specifically interact with the binding site of SH3 domains

Unexpectedly, several SH3 and EVH1 domains have been shown to interact with two clones which are not enriched in proline residues, the NP1 and NP2 variants.

NP1	Ser	Thr	Leu	**Cys**	**Asp**	**Gly**	**Tyr**	**Cys**
NP2	Asn	Pro	Met	Gly	**Asp**	**Gly**	**Tyr**	Phe

When tested on phage ELISA, both clones were demonstrated to bind with all of the SH3 and EVH1 domains tested (Figure III-43). However, NP1 and NP2 interact with each module with different relative affinities, demonstrating higher specificity for particular domains (Src, nephrocystin, Grb2, Nck1 and Abl SH3 domains, Vesl and Mena EVH1 domains). Both non proline clones react weakly with Nck2 and Crk SH3 domains, and with Evl and VASP EVH1 domains.

Panning selections on Src, nephrocystin and Nck2 SH3 domains provided interesting ligand consensus sequences. However, the CPLPPXP library was screened on several other SH3 modules, which led only to the isolation of both non proline clones (such as Nck1, Abl, ALP, $p67^{phox}$ or p53bp2 SH3 domains).

The recurrent isolation of these both clones, might mean that NP1 and NP2 were the best ligands for the target, which could have been selected from CPLPPXP library. This is unlikely, since the CPLPPXP library was designed to isolate high affinity ligands, but possible if the PPXP core motif was not the most appropriate proline motif for the target (as was observed for the nephrocystin SH3 domain).

The second hypothesis to explain the high frequency identification of NP1 and NP2, is

that the CPLPPXP library might have been contaminated with the non proline clones, and that it was then more difficult to select other variants, since NP1 and NP2 have a selective advantage, due to their shorter replication time (their are 5 amino acids shorter than proline rich clones). Lowman and Wells (1991) discussed the problem of contamination of phagemid libraries with wild-type or affinity selected variants. In some cases, the contaminants have high affinity for the target, and are amplified round after round at the expense of the desired clones. However, it happens that the contaminants are low affinity binding clones, but repetitive contamination of a library may give rise to a significant subpopulation of contaminant.

NP1 ands NP2 have been shown to interact *in vitro*, on phage ELISA, with a panel of SH3 and EVH1 domains. But it was also demonstrated that they bind to the domains, through their active binding site, where the proline rich ligands specifically interact. This was tested by competitive phage ELISA, where interaction between Src SH3 domain, and NP1 or NP2 clones was analyzed, in the presence of increasing amounts of competitor (Figure III-45). The chosen competitor was the S1 synthesized free peptide, assumed to be the best characterized Src ligand, according to its affinity constant. On phage ELISA, free S1 peptide competed with the S1, S2, S3, S4, NP1 and NP2 phage displayed peptides for SH3 binding (Figures III-44 and III-45).

This clearly demonstrated that both types of non-proline and proline rich ligands interact with Src SH3 domains at the same binding site. It might then be expected that NP1 and NP2 bind the other tested SH3 and EVH1 domains at their active binding site.

The relative binding differences observed from interaction between NP1, NP2 and individual SH3 or EVH1 domains on phage ELISA, suggest that the non proline clones bind the domains in a protein specific way. The SH3 binding site is composed of a hydrophobic groove flanked by the n-Src and RT loops, where the proline core motif anchors a ligand to the module, and of a specificity pocket, which is structurally diverse amongst the SH3 family members, and which allows extensive contact with the residues flanking the proline motif. The specificity for each particular SH3 domain is determined by this latter interaction. Therefore, the NP1 and NP2 ligands are expected to bind to SH3 domains, via interaction with all or a part of the specificity pocket.

3-2 Non proline ligands for SH3 domains

SH3 domains are known to bind proline rich ligands. However, several cases have been

reported, where SH3 domains were shown to be able to interact with non proline rich binders.

In their study of left handed polyproline II helices, Adzhubei and Sternberg (1993) showed that proline is the most preferred residue of PPII helices, and is highly advantageous for protein regions which fold as PPII helices. However, proline is not crucial for PPII helices, since within the 96 PPII helices listed by Adzhubei and Sternberg, 28 are formed without a single proline.

The cdc25 protein of the *Saccharomyces cerevisiae* which stimulates the Ras protein activity, includes a SH3 domain. Buu et al. (1995) isolated a yeast protein, the glyceraldehyde-3-phosphate dehydrogenase (GAP3DH) able to interact with the cdc25 SH3 domain. However, sequences of GAP3DH revealed no proline rich regions susceptible of interaction with SH3 domains. Nevertheless, Adzhubei and Sternberg (1993) have pointed out that GAP3DH presents one short stretch encompassing three residues which adopts a PPII helix. Buu et al. suggested that the interaction between the cdc25 SH3 domain, and the GAP3DH is mediated through this PPII region. However, they did not demonstrate it experimentally, and the binding region was not mapped.

Growth factors that activate transmembrane tyrosine kinase receptors also activate Src. The Src protein exists in repressed and activated states. SH3 and SH2 domains, and the C terminal tail have roles in regulating Src kinase activity. Phosphorylation of Tyr527 inhibits Src catalytic activity, by creating an intramolecular binding site for the SH2 domain, which locks the molecules in an inactive state (Cooper and Howell, 1993). Xu et al. (1997) crystallized human c-Src in a closed inactive conformation. They showed that a 14 residue polypeptide linker which joins the SH2 and the catalytic domains, interacts along its course with the SH3 domain, and that this interaction contributes to the stability of the closed state.

Although this linker contains only one proline (PTA**KPQTQ**GLAKDA), it adopts, from residues 249 to 253 (in bold), a polyproline type II helical conformation in complex with the recognition surface of the SH3 domain. C terminal to the Gln253, the linker is not in PPII conformation, but continues to make extensive contact with the SH3 domain.

By using an original “mirror image” phage display method, Schumacher et al. (1996) succeeded to isolate D-enantiomer non proline rich ligands against Src SH3 domain. The two D-ligands characterized with this technique against the Src SH3 domain do not resemble the

usual SH3 ligands, and are bracketed by cysteines. But since they are D-enantiomer, the sequences are not comparable with usual ligands.

More recently, Mongiovi et al. (1999) described a novel peptide-SH3 interaction. They demonstrated that the SH3 domain of Eps8, a substrate of receptor and non-receptor tyrosine kinases, interact with peptides containing the minimal consensus sequence: PXXDY.

The crystal structure of Eps8-SH3 domain revealed an intertwined dimer, in which the two anti-parallel β sheets that are characteristic for the SH3 fold, are constituted by different polypeptide chains (Kishan et al., 1997).

In Eps8 dimer, the n-Src loop from each monomer extends into the neighboring molecule. As a result, part of the classical PPII binding groove of each monomer is included in the dimer interface, and is therefore inaccessible to ligands as they are bound in other SH3 domains (Kishan et al., 1997). Kishan et al. (1997) suggested that this unique organization of the Eps8 SH3 binding site appears to be the explanation for the observation that the usual PXXP proline motif peptides are not the optimal ligands.

However, Mongiovi et al. (1999) showed that the monomer form of Eps8 SH3 domain is the one active in ligand binding, meaning that peptides with the PXXDY consensus motif bind the monomeric form of Eps8 SH3 domain. The three dimensional structure of the monomeric Eps8 SH3 domain is unknown. However, amino acid conservation at positions involved in the formation of the hydrophobic groove suggests high structural conservation between Eps8 SH3 domain, and other members of the SH3 family whose monomeric three dimensional structures are known. Thus, the difference in specificities between Eps8 and other SH3 domains, might be due to residues which are not conserved in the SH3 family, likely in the specificity pocket which is diverse amongst SH3 family members.

It is interesting to compare the non proline clones isolated in this work, NP1 and NP2, with the ligand preferences of Eps8 SH3 domain. Some similarities are evident, since the Eps8 presents a **Asp Tyr** motif, and both NP1 and NP2 contain a **Asp Gly Tyr** motif. The question as to whether these motifs are related remains opened. Could the NP1 and NP2 ligands bind Eps8 SH3 domain? The presence of the glycine between the Asp and the Tyr might disrupt the interaction, however, since Gly is the smallest naturally occurring amino acid, contact between Eps8 SH3 binding site, and NP1 or NP2 could be conceived. NP2 is the most likely from NP1 and NP2 ligands, to be able to interact with Eps8 SH3 domain, due to the presence of the proline three positions before the Asp.

Eps8 motif		**Pro**	Xxx	Xxx	**Asp**	**Tyr**		
NP1	Ser	Thr	Leu	***Cys***	**Asp**	**Gly**	**Tyr**	***Cys***
NP2	Asn	**Pro**	Met	Gly	**Asp**	**Gly**	**Tyr**	Phe

The Eps8 consensus sequence proves that is possible for a ligand to bind SH3 domains without the commonly conserved Arg responsible of the classification of SH3 ligands. It is not known if the aspartic acid found in NP1, NP2 and the Eps8 motif, can form a salt bridge with a conserved residue of the SH3 domains.

Taken together, these data proved that non proline clones can interact with SH3 domains, even if they represent a minority of the isolated ligands. It has also been demonstrated that proline-free sequences are able to form PPII helices (Adzhubei and Sternberg, 1993). This suggests that SH3 ligand preferences require the presence of the PPII helices, and that prolines are found in the SH3 ligands, since they are advantageous to adopt a PPII conformation, and not for their particular side-chain interaction with the modules. This is supported by the work of Nguyen et al. (1998), who showed that SH3 domains broadly interact with ligands containing non natural amide N-substituted residues, instead of the usual prolines. They suggested that proline might be specifically selected because it is the only endogenous N-substituted amino acid. In addition, Pisabarro and Serrano (1996) showed that all proline residues of a SH3 ligand can be substituted by other amino acids, as long as a replacement, or a compensating mutation is introduced elsewhere. They suggested that there is no essential proline residue in the interaction of a peptide with a SH3 domain, and that proline rich ligands are selected due to the reduction of the entropic cost they provide, when a region has to adopt a PPII helix.

4- Transfection of the isolated peptides - *in vivo* interaction

The ligands isolated with the CPLPPXP library against Src and nephrocystin SH3 domains were shown to interact *in vitro* with their target with very a high affinity (K_D in the nanomolar range). Such high affinity ligands might be able to compete *in vivo* with natural SH3 ligands, and thus, to be used to elaborate cellular models where the signal transduction pathways could be tightly controlled and regulated. Such models would provide information allowing a better understanding of the signal transduction pathways, and of the roles of SH3 domains. Before elaborating cellular models, it is crucial to ensure that the CPLPPXP isolated peptides are able to bind *in vivo* to the SH3 domains.

The Tat fusion system developed by Dowdy's group (Nagahara et al., 1998) was used to

deliver the best characterized ligands against Src and nephrocystin SH3 domains within NIH 3T3 fibroblast cells. The Tat-S4 peptide, specific for Src and nephrocystin SH3 domains, was used for the *in vivo* analyses. Tat fusion peptides were fused to the fluorochrom FITC, to facilitate their detection within the cells. It was shown by FACS analysis that 77% of the cells were efficiently transduced with the FITC-peptide, after one hour incubation (Figure III-55/A-B). However, no clear concentration dependence was detected (Figure III-55/C), in contrast to the data reported by Dowdy (Nagahara et al., 1998; Vocero-Akbani et al., 1999).

Confocal microscopic analysis of cells transfected with the Tat fusion peptides provided interesting observations (Figure III-56). The FITC-Tat-peptides were detected within the cytoplasm, but not within the nucleus. However, since the Tat protein sequence contains a nuclear localization signal, the Tat fusion peptides were expected to be located in both cytoplasmic and nuclear compartments, as observed in Dowdy's work (Nagahara et al., 1998). The retention of the Tat fusion peptide within the cytoplasm, suggest that the peptides are maintained there, likely by interaction with the cytoplasmic SH3 containing proteins.

The FITC-Tat-peptides were still detected within the cytoplasm of the cells after a few days of culture. No trace of FITC labeled Tat-peptides was detected in the nucleus. However, Tat fusion protein transfection is transient, and the fusion proteins are able to cross the cell membrane in both directions. Therefore, the observation that Tat-peptides are maintained within the cytoplasm also supports the hypothesis that peptides are retained there through an *in vivo* interaction with other proteins, which are likely to be Src protein, or relatives.

The confocal microscopy analysis shows different localization of the FITC-Tat-peptides in healthy or dead cells (Figure III-56). In healthy cells, the fusion peptide was found ubiquitously in the cytoplasm, whereas, in dead cells, the Tat fusion peptide is shown to be exposed at the border of the cell, close to the cell membrane.

It is unclear why the Tat fusion peptides are directed at this location. Src family kinases are proteins associated with both plasma and intracellular membranes. It is however, unknown whether after cell death, the Src proteins are directed or not to the cell plasma membrane, where they would trap the Tat-fusion peptides. Further investigations would be necessary to answer these questions.

V - CONCLUSION

A dedicated biased cosmix-plexing® phage-display library, CPLPPXP, displaying the partially hypervariable sequence X_5PPXPX_4, has been created, in order to isolate high affinity ligands for specific targets, namely from the SH3- and EVH1-domain families.

The initial use of the cosmix-plexing® library on SH3 and EVH1 domains, showed that long consensus sequences for specifically binding ligands can be readily obtained.

During this work, it was noticed that the cosmix-plexing® recombination strategy offers advantages that vary from one target to another: In the case of Src SH3 domain, the cosmix-plexing® recombination of a population selected after panning with the primary library, permitted the confirmation of the importance of the S1 ligand in Src interactions; in other cases, it allowed the isolation of novel ligands for a particular target, for instance nephrocystin SH3 domain and Vesl EVH1 domain; in general, the cosmix-plexing® strategy led to less ambiguities and longer consensus sequences for each target used (an exception was the Src-SH3 domain, since exclusively one clone was enriched); in the case of Nck2 SH3 domain, only the post-recombination cosmix-plexing® library allowed characterization of Nck2 ligands, leading to the conception of a consensus sequence. Affinity selection with the primary library remained unsuccessful, and did not result in quantitative enrichment of clones.

The use of the cosmix-plexing® library clearly allowed isolation of SH3 ligands with affinities 100- to 1000-fold higher than ligands previously described, according to BIAcore analysis. On the basis of this data, the cosmix-plexing® recombination system can be considered as the most powerful technique for the isolation of high affinity ligands.

From the initial experiments to deliver the cosmix-plexing® SH3 ligands *in vivo* into NIH-3T3 fibroblasts, it seems that they specifically interact with their target, since they do not enter the nucleus, but are maintained within the cytoplasm. Moreover they remain a few days within the cytoplasm despite the fact that they are freely permeable to the cell membranes.

We conclude that the method is convenient to characterize high affinity ligands, which might bind strongly enough to the target, to compete *in vivo* with natural ligands.

VI - ABSTRACT

Phage display is a technique by which variant peptides can be displayed as fusion proteins to a coat protein on the surface of filamentous phage particles. The development of a novel phage display strategy, termed cosmix-plexing®, which allows the isolation of high affinity ligands against a particular target, is described here.

The work focused on characterization of high affinity ligands to SH3 domains, modules found in a variety of unrelated proteins, many of which are involved in signal transduction, indicating their important role in protein-protein interactions. SH3 domains show a high preference for proline-rich ligands, and therefore, a dedicated biased cosmix-plexing® library was designed, containing three prolines at fixed positions (X_5PPXPX_4). EVH1 domains, protein-protein interaction modules implicated in the spatial control of actin assembly, also mediate the interaction with their partners through a proline rich motif, and were therefore also considered as a target of interest for this library.

Cosmix-plexing® allows recombination within hypervariable DNA regions, due to a type IIs enzyme, which binds at a defined sequenced, but cleaves at a site a fixed distance away, i.e. in a region with no demands on sequence specificity. This technique permits to increase the diversity of an initial library, however, its main strength is to allow the reassortment of the right and left sections of hypervariable domains of a preselected population, which showed affinity for a particular target, and thus, to optimize the binding capacities of the variants, for this target.

High affinity ligands to SH3 and EVH1 domains, were isolated from the cosmix-plexing® library. BIAcore analyses were performed on SH3 ligands, which demonstrated that these cosmix-plexing® ligands show dissociation constants in the low nanomolar range, i.e. they interact with SH3 domains with an affinity approximately 1000 fold higher than previously described SH3 binders.

Such high affinity ligands allow the development of tools to investigate the function of SH3 domain containing proteins in the signal transduction cascade. One of these is the elaboration of cellular models, where SH3 natural ligands might be competed away *in vivo,* and where the signal transduction pathways could be tightly controlled and regulated. Preliminary experiments to deliver the cosmix-plexing® SH3 peptides into living cells in cell culture, using the Tat-fusion protein transduction system, allow optimism regarding these cellular models.

Zusammenfassung

Phage display stellt eine Technik dar, mit der verschiedene Peptide zusammen mit einem Oberflächenprotein des filamentösen Phagenpartikels als Fusionsproteine exprimiert werden können. Hier wird die Entwicklung einer neuen Phage display Strategie beschrieben, die als cosmix-plexing® bezeichnet wird. Sie erlaubt die Isolation hoch affiner Liganden.

Die Arbeit konzentriert sich auf die Charakterisierung von Liganden mit hoher Affinität gegen SH3-Domänen. SH3-Domänen finden sich in vielen nicht verwandten Proteinen und vermitteln häufig Signaltranduktion, was ihre wichtige Rolle bei Protein-Protein-Wechselwirkungen verdeutlicht. Es wurde gezeigt, dass SH3-Domänen vorwiegend an Prolin reiche Liganden binden. Deshalb wurde eine Ziel gerichtete cosmix-plexing® Bibliothek generiert, die drei Prolinreste an definierten Positionen beinhaltet (X_5PPXPX_4).

EVH1 Domänen, die an der Kontrolle der Aktin-Zusammenlagerung beteiligt sind, vermitteln die Protein-Wechselwirkung mit ihren Liganden ebenfalls durch ein Prolin reiches Motiv. Deshalb kamen sie als interessante Domäne für diese Bibliothek in Betracht.

Cosmix-plexing® erlaubt die Rekombination in hypervariablen DNA Regionen. Dazu ist ein Restriktionsenzym vom Typ IIs erforderlich, das an eine definierte DNA Sequenz bindet, aber in einem festgelegten Abstand dazu schneidet: Das heißt, die Bindung erfolgt in einer Region ohne Sequenzspezifität. Diese Technik erlaubt es die Diversität der Ausgangsbibliothek zu steigern. Der Hauptvorteil dieser Methode liegt aber darin, dass sie die Rekombination der rechten und linken Spaltstücke der hypervariablen Domänen einer vorselektierten Population erlaubt, die bereits eine Affinität für ihre bestimmte Zieldomäne besitzt. Dadurch wird die Affinität der Varianten für ihr Zielprotein optimiert.

Liganden mit hoher Affinität für SH3 und EVH1 Domänen konnten mit einer cosmix-plexing® Bibliothek isoliert werden. BIAcore Analysen für diese isolierten SH3 Liganden zeigten Dissoziationskonstanten im unteren Nanomolar-Bereich, d.h. sie besitzen eine etwa tausendfach höhere Affinität zu SH3-Domänen, als die bisher beschriebenen SH3-Liganden

Solche hoch affinen Liganden stellen Hilfsmittel dar, um die Funktion von SH3 Domänen enthaltenden Proteinen in Signaltransduktions-Kaskaden zu untersuchen. Zelluläre Modellsysteme könnten ausgearbeitet werden, in denen natürliche SH3-Liganden *in vivo* kompetitiert und in denen Signaltransduktionswege streng kontrolliert und reguliert werden. Vorläufige Experimente, in denen cosmix-plexing® SH3 Peptide mit Hilfe des Tat-fusion protein transduction systems in lebende Zellen eingebracht wurden, lassen auf neue Einblicke in solchen zellulären Modellen hoffen.

VII - REFERENCES

Adzhubei, A.A., Sternberg, M.J.E. (1993). Left-handed polyproline II helices commonly occur on globular proteins. *J. Mol. Biol.* 229, 472-493.

Albini, A., Barillari, G., Benelli, R. Gallo, R.C:, Ensoli B. (1995). Angiogenic properties of human immunodeficiency virus type 1 Tat protein. *Proc. Natl. Acad. Sci. USA.* 92, 4838-4842.

Alexandropoulos, K., Cheng, G., Baltimore, D. (1995). Proline-rich sequences that bind to Src homology 3 domains with individuals specificities. *Proc. Natl. Acad. Sci. USA.* 92, 3110-3114.

Andersson, P., Goldfarb, M.P., Weinberg, R.A. (1979). A defined subgenomic fragment of in vitro synthesised Moloney sarcoma virus DNA can induce cell transformation upon transfection. *Cell.* 16, 63-75.

Antignac, C., Arduy, C.H., Beckmann, J.S., Benessy, F., Gros, F., Medhioub, M., Hildebrandt, F. (1993). A gene for familial juvenile nephronophthisis (recessive medullary cystic kidney disease) maps to chromosome 2p. *Nature Genet.* 3, 342-345.

Anton, I.M., Lu, W., Mayer, B.J., Ramesh, N., Geha, R.S. (1998). The Wiscott-aldrich syndrome protein-interacting protein (WIP) binds to the adaptor protein Nck. *J. Biol. Chem.* 273 (33), 20992-20995.

Armstrong, N., Adey, N.B., McConnell, S.J., Kay, B.K. (1996). Vectors for phage display. In *Phage display of peptides and proteins: a laboratory manual*. Kay, B.K., Winter J., McCafferty, J. Eds. Academic press, Inc, San Diego. pp 35-53.

Aszdoti, A., Pfeifer, A., Ahmad, M., Glauner, M., Zhou, X.-H., Ny, L., Andersson, K.-E., Kehrel, B., Offermans, S., Fässler, R. (1999). The vasodilator-stimulated phosphoprotein (VASP) is involved in cGMP- and cAMP-mediated inhibition of agonist-induced platelet aggregation but is dispensable for smooth muscle function. *EMBO. J.* 18, 37-48.

Barfod, E.T., Zheng, Y., Kuang, W.-J., Hart, M.J., Evans, T., Cerione, R.A., Ashkenazi, A. (1993). Cloning and expression of a human cdc42 GTPase-activating protein reveals a functional SH3-binding domain. *J. Biol. Chem.* 268, 26059-26062.

Beck, E., Zink, B. (1981). Nucleotide sequence and genome organization of filamentous bacteriophage fl and fd. *Gene*. 16, 35-58.

Birnoim, H.C., Doly, J. (1979). A rapid alkaline extraction procedure for screening re-combinant plasmid DNA. *Nucl. Acids Res*. 7, 1513-1522.

Bokoch, G.M., Wang, Y., Bohl, B.P., Sells, M.A., Quilliams, L.A., Knaus, U.G. (1996). Interaction of the Nck adapter protein with p21-activated kinase (PAK1). *J. Biol. Chem*. 271, 25746-25749.

Bradford, M.M. (1976). A rapid and sensitive method for the quantification of microgram quantities protein utilizing the principle of protein-dye binding. *Anal. Biochem*. 72, 248-254.

Brakeman, P.R., Lanahan, A.A., O'Brien, R., Roche, K., Barnes, C.A., Huganir, R.L., Worley, P.F. (1997). Homer: a protein that selectively binds metabotropic glutamate receptors. *Nature*. 386, 284-288.

Brindle, N.P.J., Holt, M.R., Davies, J.E., Price, C.J., Critchley, D.R (1996). The focal-adhesion vasodilator-stimulated phosphoprotein (VASP) binds to the proline-rich domain in vinculin. *Biochem. J*. 318, 753-757.

Brown, M., T., Cooper, J.A. (1996). Regulation, substrates and functions of Src. *Biochimica et Biophysica Acta*. 1287, 121-149.

Buday, L. (1999). Membrane-targeting of signaling molecules by SH2/SH3 domain-containing adaptor proteins. *Biochimica et Biophysica Acta*. 1422, 187-204.

Bunnell, S.C., Henry, P.A., Kolluri, R., Kirchhausen, T., Rickles, R.J., Berg, L.J. (1996). Identification of Itk/Tsk Src homology 3 domain ligands. *J. Biol. Chem*. 271 (41), 25646-25656.

Buu, A., Garreau, H., Jacquet, M. (1995). The glyceraldehyde-3.phosphate dehydrogenase binds in vitro to the SH3 domain of *Saccharomyces cerevisiae* Cdc25p. *Crit. Rev. Acad. Sci. Paris Sci. Vie/Life Sci.* 318, 665-669.

Carl, U.D., Pollmann, M., Orr, E., Gertler, F.B., Chakraborty, T., Wehland, J. (1999). Aromatic and basic residues within the EVH1 domain of VASP specify its interaction with proline-rich ligands. *Curr Biol*. 9 (13), 715-718.

Chakraborty, T., Ebel, F., Domann, E. Niebuhr, K., Gerstel, B., Pistor, S., Temm-Grove, C.J., Jockusch, B.M., Reinhard, M., Walter, U., Wehland, J. (1995). A focal adhesion factor directly linking intracellularly motile *Listeria monocytogenes* and *Listeria ivanovii* to the actin-based cytoskeleton of mammalian cells. *EMBO. J.* 14, 1314-1321.

Cheadle, C., Ivashchenko, Y., South, V., Searfoss, G.H., French, S., Howk, R., Ricca, G.A., Jaye, M. (1994). Identification of a Src SH3 domain binding motif by screening a random phage display library. J. *Biol. Chem*. 269, 24034-24039.

Chen, J.K., Lane, W.S., Brauer, A.W., Tanaka, A., Schreiber, S.L. (1993). Biased combinatorial libraries: novel ligands for the SH3 domain of Phosphatidylinositol 3-Kinase. *J. Am. Chem. Soc.* 115, 12591-12592.

Chou, P.Y., Fasman, G.D. (1978). Prediction of the secondary structure of proteins from their amino acid sequence. *Advan. Enzymol.* 47, 145-148.

Chou, M., Hanafusa, H. (1995). A novel ligand for SH3 domains. *J. Biol. Chem*. 270, 7359-7364.

Choudhury, G.G., Marra, F., Abboud, H.E. (1996). Thrombin stimulates association of src homology domain containing adaptor protein Nck with pp125FAK. *Am. J. Physiol*. 270, F295-F300.

Cicchetti, P., Mayer, B.J., Thiel, G., Baltimore, D. (1992). Identification of a protein that binds to the SH3 region of Abl and is similar to Bcr and GAP-rho. *Science*. 257, 803-806.

Clackson, T., Hoogenboom, H.R., Griffiths, A.D., Winter, G. (1991). Making antibody fragments using phage display libraries. *Nature* (London). 352, 624-628.

Clackson, T., Wells, J.A. (1994). *In vitro* selection from protein and peptide libraries. *Trends Biotechnol*. 12, 173-184.

Cline J., Braman, J.C., Hogrefe, H.H. (1996). PCR fidelity of Pfu DNA polymerase and other thermostable DNA polymerases. *Nucleic Acids Res*. 24 (18), 3546-3551.

Cohen, A.H., Hoyer, J.R. (1986). Nephronophthisis: a primary tubular basement membrane defect. *Lab. Invest*. 55, 564-572.

Cohen, G.B., Ren, R., Baltimore, D. (1995). Modular binding domains in signal transduction proteins. *Cell.* 80, 237-248.

Collins, J. (1981). Instability of palyndromic DNA in Escherichia coli. Cold Spring Harbor Symp. *Quant. Biol*. 45 Pt1, 409-416.

Collins, J., Volckaert, G., Nevers, P. (1982). Precise and nearly-precise excision of the symmetrical inverted repeats of Tn5; common features of recA-independent deletion events in Escherichia coli. *Gene*. 19, 139-146.

Collins, J., Szardenings, M., Maywald, F., Blöcker, H., Frank, R., Hecht, H.J., Vasel, B., Schomburg, D., Fink, E., Fritz, H. (1990). Human leukocyte elastase inhibitors: designed variants of human pancreatic secretory trypsin inhibitor (hPSTI). *Biol. Chem*. Hoppe-Seyler 371, 29-36.

Collins, J. (1997). Phage display. In *Ann. Reports in Combinatorial Chem. & Mol. Diversity*. W.H. Moos et al (Eds.), ESCOM. Leiden. Vol. 1, 210-250.

Collins, J. Röttgen, P.(1997a). EP 9800533 (pending) based on EP 97101539.1, (priority 6.02.97). Cosmix-plexing® patent.

Collins, J., Röttgen, P. (1997b). Evolutive phage-display – a method as basis for a company. *GBF, Scientific annual report*, pp 19-28.

Cooper, J.A., Howell, B. (1993). The when and how of Src regulation. *Cell*. 73, 1051-1054.

Copeland, N.G., Cooper, G.M. (1979). Transfection by exogenous and endogenous murine retrovirus DNAs. *Cell*. 16, 347-56.

Courtney, B.C., Williams, K.C., Schlager, J.J. (1995). A phage display vector with improved stability, applicability and ease of manipulation. *Gene*. 165 (1), 139-140.

Cowburn, D., Zheng, J., Xu, Q., Barany, G. (1995). Enhanced affinities and specificities of consolidated ligands for the SH3 and SH2 domains of Abelson (Abl) protein tyrosine kinase. *J. Biol. Chem*. 270 (45), 26738-2674.

Cowburn, D., Kuriyan, J. (1996). SH2, SH3 and PH domains. In *Signal Transduction*. Carl-Henrik Heldin and Mary Purton eds. Published by Chapman and Hall. pp127- 141.

Crameri, R., Suter, M. (1993). Display of biologically active proteins on the surface of filamentous phages: a cDNA cloning system for selection of functional gene products linked to the genetic information responsible for their production. *Gene*. 137, 69-75.

Cunningham, B.C., Lowe, D.G., Li, B., Bennett, B.D., Wells, J.A. (1994). Production of an atrial natriuretic peptide variant that is specific for type A receptor. *EMBO J*. 13, 2508-2515.

Dennis, M.S., Lazarus, R.A. (1994). Kunitz domain inhibitors of tissue factor-factor VIIa. II. Potent and specific inhibitors by competitive phage selection. *J. Biol. Chem.* 269, 22137-22144.

Diacumakos, E-G. (1973). Methods for micromanipulation of human somatic cells in culture. *Methods Cell. Biol.* 7, 287-311.

Dedman, J.R., Kaetzel, M.A., Chan, H.C., Nelson, D.J., Jamieson, J.G.A. (1993). Selection of target biological modifiers from a bacteriophage library of random peptides: the identification of novel calmodulin regulatory peptides. *J. Biol. Chem*. 268, 23025-23030.

De La Cruz, V., Laa, A., McCutchan, T. (1988). Immunogenicity and epitope mapping of foreign sequences via a genetically engineered filamentous phage. *J. Biol. Chem*. 263, 4318-4322.

Dennis, M.S., Lazarus, R.A. (1994). Kunitz domain inhibitors of tissue-factor factor VIIa. I. Potent inhibitors selected from libraries by phage display. *J. Biol. Chem*. 269, 22129-22136.

Derossi, D., Calvet, S., Trembleau, A., Bruissen, A., Chassaing, G. (1996). A cell internalization of the third helix of the *Antennapedia* homeodomain is receptor-independent. *J. Biol. Chem*. 271, 18188-18193.

Dooley, C.T:, Houghten, R.A. (1993). The use of positionally scanning synthetic combinatorial libraries for the rapid determination of opioid receptor ligands. *Life Sci.* 52, 1509-1517.

Dower, W.J., Miller, J.F., Ragsdale, C.W. (1988). High efficiency transformation of *E. coli* by high voltage electroporation. *Nucleic Acids Res*. 16, 6127-6145.

Duan, Y., Laursen, R.A. (1993). Protease substrate specificity mapping using membrane-bound peptides. *Anal. Biochem*. 216: 431-438.

Dube, S. (1997). Top ten facts for DNA electrophoresis and troubleshooting DNA agarose gel electrophoresis. *Focus (technical tips publication of Gibco BRL life technologies).* 19 (3), 65-66. Or accessible via internet: http://www2.lifetech.com/focus_page.htlm

Duenas, M., Borrebaerck, C.A.K. (1994). Clonal selection and amplification of phage displayed antibodies by linking antigen recognition and phage replication. *Bio/Technology*. 12, 999-1002.

Eichholtz, T., De Bont, D.B., De Widt, J., Liskamp, R.M., Ploegh, H.L. (1993). A myristoylated pseudosubstrate peptide, a novel protein kinase C inhibitor. *J. Biol. Chem*. 268, 1982-1986.

Erpel, T., Courtneidge, S.A. (1995). Src family protein tyrosine kinases and cellular signal transduction pathways. *Curr. Opin. Cell. Biol*. 7, 176-182.

Fägerstam, L., Frostell, A., Karlsson, R., Kullman, M., Larsson, A., Malmqvist, M., Butt, H. (1991). Detection of antigen-antibody interactions by surface plasmon resonance. Application to epitope mapping. *J. Mol. Recognition*. 3, 208-214.

Fawell, S., Seery, J., Daikh, Y., Moore, C., Chen, L.L., Pepinsky, B., Barsoum, J. (1994).Tat-mediated delivery of heterologous proteins into cells. *Proc. Natl. Acad. Sci. USA*. 91, 664-668.

Fedorov, A.A., Fedorov, E., Gertler, F., Almo, S.C. (1999). Structure of EVH1, a novel proline-rich ligand binding module involved in cytoskeletal dynamics and neural function. *Nature structural biology*. 6 (7), 661-665.

Feller, S.M., Posern, G., Voss, J., Kardinal, C., Sakkab, D., Zheng, J., Knudsen, S.S. (1998). Physiological signals and oncogenesis mediated through Crk family adapter proteins. *J. Cellular Physiology.* 177, 535-552.

Feng, S., Chen, J.K., Yu, H., Simon, J.A., Schreiber, S.L. (1994). Two binding orientations for peptides to the Src SH3 domain: development of a general model for the SH3-ligand interactions. *Science*. 266, 1241-1247.

Feng. S., Kasahara, C., Rickles, R.J., Schreiber, S.L. (1995). Specific interactions outside the proline-rich core of two classes of Src Homology 3 ligands. *Proc. Natl. Acad. Sci. USA*. 92, 12408-12415.

Fields, G.B., Noble, R.L. (1990). Solid phase synthesis utilizing 9-fluorenylmethoxycarbonyl amino acids. *Int. J. Peptide Proteins Res*. 35, 161-214.

Fields, S., Song, O.-K. (1989). A novel genetic system to detect protein-protein interactions. *Nature*. 340, 145-246.

Finan. P.M., Soames, C.J., Wilson, L., Nelson, D.L., Stewart, D.M., Truong, O., Hsuan, J.J., Kellies, S. (1996). Identification of regions of the Wiskott-Aldrich syndrome protein responsible for association with selected Src homology 3 domains. *J. Biol. Chem*. 271, 26291-26295.

Fisch, I., Kontermann, R.E., Finnern, R., Hartley, O., Soler-Gonzales, A.S., Griffiths, A.D., Winter, G. (1996). A strategy of exon shuffling for making large peptide repertoires displayed on filamentous phage. *Proc. Natl. Acad. Sci. USA*. 93, 7761-7766.

Folgori, A., Tafi. R., Meola. A., Felici. F., Galfre. G., Cortese. R., Monaci. P., Nicosia. A. (1994). A general strategy to identify mimotopes of pathological antigens using only random peptide libraries and human sera. *EMBO J*. 13, 2236-2243.

Francisco, J.A., Campbell, R., Iverson, B.L., Georgiou, G. (1993). Production and fluorescence-activated cell sorting of *Escherichia coli* expressing a functional antibody fragment on the external surface. *Proc. Natl. Acad. Sci. USA*. 90, 1044-10448.

Frank, R. (1992). Spot-synthesis: an easy technique for positionally addressable, parallel chemical synthesis on a membrane support. *Tetrahedron*. 48 (42), 9217-9232.

Frank, R., Kieß, M., Lahmann, H., Behn, C., Gausepohl, H. (1995). Combinatorial synthesis on membrane supports by the SPOT technique. In *peptides 1994* (Maia, H., Eds.), ESCOM, Leiden. 479-480.

Frankel, A.D., Pabo, C.O. (1988). Cellular uptake of the tat protein from human immunodeficiency virus. Cell. 55: 1189-1193.

Fuchs, P., Weichel, W., Dübel, S., Breitling, F., Little, M. (1996). Separation of *E.coli* expressing functional cell-wall bound antibody fragments by FACS. *Immunotechnology*. 2, 97-102.

Freeman, N.L., Chen, Z., Horenstein, J., Weber, A., Field, J. (1995). An actin monomer binding activity localizes to the carboxyl-terminal half of the *Saccharomyces cerevisiae* cyclase-associated protein. *J. Biol. Chem*. 270 (10), 5680-5685.

Freeman, N.L., Lila, T., Mintzer, K.A., Chen, Z., Pahk, A.J., Ren, R., Drubin, D.G., Field, J. (1996). A conserved proline-rich region of the Saccharomyces cerevisiae cyclase-associated protein binds SH3 domains and modulates cytoskeletal localization. *Mol. Cell. Biol*. 16, 548-556.

Fulford, W., Model, P. (1984). Gene X of bacteriophage fl is required for phage DNA synthesis: Mutagenesis of in-frame overlapping genes. *J. Mol. Biol*. 178, 137-153.

Galisteo, M.L., Chernoff, J., Su, Y.-C., Skolnik, EY., Schlessinger, J. (1996). The adapter protein Nck links receptor tyrosine kinases with the serine-threonine kinase Pak1. *J. Biol. Chem*. 271, 20997-21000.

Gallop, M.A., Barrett, R.W., Dower, W.J., Fodor, S.P.A., Gordon, E.M. (1994). Applications of combinatorial technologies to drug discovery. 1. Background and peptide combinatorial libraries. *Journal of Medicinal Chemistry*. 37 (9), 1233-1251.

Gee, S.H., Sekely, S.A., Lombardo, C., Kurakin, A.,Froehner, S.C., Kay, B.K. (1998). Cyclic peptides as non-carboxyl-terminal ligands of syntrophin PDZ domains. *J. Biol. Chem*. 273 (34), 21980-21987.

Geoffroy, F., Sodoyer, R., Aujame, L. (1994). A new phage display system to construct multicombinatorial libraries of very large antibody repertoires. *Gene*. 151, 109-113.

Georgiou, G., Stathopoulos, C., Daugherty, P.S., Nayak, A.R., Iverson, B.L., Curtiss III, R. (1997). Display of heterologous proteins on the surface of microorganisms: from the screening of combinatorial libraries to live recombinant vaccines. *Nature Biotechnology*. 15, 29-34.

Gertler, F.B., Niebuhr, K., Reinhard, M., Wehland, J., Soriano, P. (1996). Mena, a relative of VASP and *Drosophila* Enabled, is implicated in the control of microfilament dynamics. *Cell*. 87, 227-239.

Geysen, H.M., Rodda, S.J., Mason, T.J. (1986). *A priori* delineation of a peptide which mimics a discontinuous antigenic determinant. *Mol. Immunol*. 23, 709-715.

Gieselmann, R., Mann, K. (1992). ASP-56, a new actin sequestering protein from pig platelets with homology to CAP, an adenylate cyclase-associated protein from yeast. *FEBS Lett*. 298, 149-153.

Gram, H., Marconi, L.A., Barbas, C.F. III, Collet, T.A., Lerner, R.A., Kang, A.S. (1992). *In vitro* selection and affinity maturation of antibodies from a naive combinatorial immunoglobulin library. *Proc. Natl. Acad. Sci. USA*. 89, 3576-3580.

Gramatikoff, K., Georgiev, O., Schaffner, W. (1994). Direct interaction rescue, a novel filamentous phage technique to study protein-protein interactions. *Nucl. Acids. Res*. 22 (25), 5761-5762.

Green, M., Loewenstein, P.M. (1988). Autonomous functional domains of chemically synthesised human immunodeficiency virus tat trans-activator protein. *Cell*. 55, 1179-1188.

Greenwood, J., Willis, A., Perham, R. (1991). Multiple display of foreign peptides on a filamentous bacteriophage: peptides from *Plasmodium falciparum* circumsporozoite protein as antigens. *J. Mol. Biol*. 220, 821-827.

Griffiths, A.D., Williams, S.C., Hartley, O., Tomlinson, I.M., Waterhouse, P., Crosby, W.L., Kontermann, R.E., Jones, P.T., Low, N.M., Allison, T.J., Prospero, T.D., Hoogenboom, H.R., Nissim, A., Cox, J.P.L., Harrison, J.L., Zaccolo, M., Gherardi, E., Winter, G. (1994). Isolation of high affinity human antibodies directly from large synthetic repertoires. *EMBO J*. 13, 3245-3260.

Halbrügge, M., Walter, U. (1989). Purification of a vasodilator-regulated phosphoprotein from human platelets. *Eur. J. Biochem*. 185, 41-50.

Hanes, J., Plückthun, A. (1999). *In vitro* selection methods for screening of peptide and protein libraries. In *Combinatorial Chemistry in Biology*, Famulok, M., Winnacker, R.-L-, Wong, C.-H. (Eds.). 107-122.

Hart, C.P., Martin, J.E., Reed, M.A., Keval, A.A., Pustelnik, M.J., Northrop, J.P., Patel, D.V., Grove, J.R (1999). Potent inhibitory ligands of the Grb2 SH2 domain from recombinant peptide libraries. *Cell. Signal*. 11(6), 453-464.

Hawiger, J. (1997). Cellular import of functional peptides to block intracellular signaling. *Curr. Opin. Immunol*. 9, 189-194.

Hawkins, R.E., Russell, S.J., Winter, G. (1992). Selection of phage antibodies by binding affinity mimicking affinity maturation. *J. Mol. Biol*. 226, 889-896.

Henkel, T:, Zabel, U., Van Zee, K., Muller, J.M., Fanning, E., Baeurerle, P.A. (1992). Intramolecular masking of the nuclear location signal and dimerization domain in the precursor for the p50 NF-kappa B subunit. *Cell*. 68, 1121-1133.

Heukeshoven, J., Dernick, R. (1988). Improved silver staining procedure for fast staining in phastsystem development unit; I. staining of sodium dodecyl sulfate gels. *Electrophoresis*. 9, 28-32.

Hildebrandt, F., Otto, E., Rensing, C., Nothwang, H.G., Vollmer, M., Adolphs, J., Hanusch, H., Brandis, M. (1997a). A novel gene encoding an SH3 domain protein is mutated in nephronophthisis type 1. *Nature Genetics*. 17, 149-153.

Hildebrandt, F., Strahm, B., Nothwang, H.-G., Gretz, N., Schnieders, B., Singh-Sawhney, I., Kutt, R., Vollmer, M., Brandis, M., and members of the APN study group. (1997b). Molecular genetic identification of families with juvenile nephronophthisis type 1: rate of progression to renal failure. *Kidney International*. 51, 261-269.

Hildebrandt, F. (1998). Identification of a gene for nephronophthisis. *Nephrol. Dial. Transplant.* 13, 1334-1336.

Hill, D.F., Petersen, G.B. (1982). Nucleotide sequence of the bacteriophage fl DNA. *J. Virol.* 44, 32-46.

Hogrefe, H.H., Amberg, J.R., Hay, B.N., Sorge, J.A., Shopes, B. (1993). Cloning in a bacteriophage lambda vector for the display of binding proteins on filamentous phage. *Gene*. 137, 85-91.

Holliger, P., Riechmann, L. (1997). A conserved infection pathway for filamentous bacteriophages is suggested by the structure of the membrane penetration domain of the minor coat protein g3p from phage fd. *Structure*. 5, 265-275.

Hoogenboom, H:, Griffiths, A., Johnson, K., Chisswell, D., Hudson, P., Winter, G. (1991). Multisubunit proteins on the surfaces of filamentous phage: methodologies for displaying antibody (Fab) heavy and light chains. *Nucleic Acids Res*. 19, 4133-4137.

Hoogenboom, H., Winter, G. (1992). By-passing immunization. Human antibodies from synthetic repertoires of germline VH gene segments rearranged in vitro. *J. Mol. Biol*. 227, 381-388.

Houshmand, H., Fröman, G., Magnusson, G. (1999). Use of bacteriophage T7 displayed peptides for determination of monoclonal antibody specificity and biosensor analysis of the binding reaction. *Analytical Biochemistry*. 268, 363-370.

Hu, Q., Milfay, D., Williams, L.T. (1995). Binding of Nck to Sos and activation of ras-dependent gene expression. *Mol. Cell Biol*. 15, 1169-1174.

Huebner, K., Kastury, K., Druck, T., Salcini, A.E., Lanfrancone, L., Pelicci, G., Lowenstein, E., Li, W., Park, S.-H., Cannizzaro, L., Pelicci, P.G., Schlesinger, J. (1994). Chromosome locations of genes encoding human signal transduction adaptor proteins, Nck (NCK), Shc (SHC1), and Grb2 (GRB2). *Genomics*. 22, 281-287.

Ishihara, H., Hara, T., Aramaki, Y., Tsuchiya, S., Hosoi, K. (1990). Preparation of asialofetuin-labeled liposomes with encapsulated human interferon-gamma and their uptake by isolated rat hepatocytes. *Pharm. Res*. 7, 542-546.

Jespers, L.S., Messens, J.H., Keyser, A.D., Eeckhout, D., Brande, I.V.D., Gansemans, Y.G., Lauwereys, M.J., Viasuk, G.P., Stassens, P.E. (1995). Surface expression and ligand-based selection of cDNA fused to the filamentous phage gene VI. *Biotechnology*. 13, 378-382.

Joliot, A., Pernelle, C., Deagostini-Bazin, H., Prochiantz, A. (1991a). *Antennapedia* homeobox peptide regulates neural morphogenesis. *Proc. Natl. Acad. Sci. USA*. 88, 1864-1868.

Joliot, A., Triller, A., Volovitch, M., Pernelle, C., Prochiantz, A. (1991b). Alpha-2, 8-polysialic acid is the neuronal surface receptor of *Antennapedia* homeobox peptide. *New Biol*. 3, 1121-1134.

Jönsson, U., Malmqvist, M. (1992). Real time biospecific interaction analysis. The integration of surface plasmon resonance detection, general biospecific interface chemistry and microfluidics into one analytical system. In *Advances in Biosensors*. Edited by F. Turner, London: JAI Press. 2, 291-336.

Kang, A.S., Barbas III, C.F., Janda, K.D., Benkovic, S.J., Lerner, R.A. (1991a). Linkage of recognition and replication functions by assembling combinatorial antibody Fab libraries along with phage surface. *Proc. Natl. Acad. Sci. USA*. 88, 4363-4366.

Kang, A.S., Jones, T.M., Burton, D.R. (1991b). Antibody redesign by chain shuffling from random combinatorial immunoglobulin libraries. *Proc. Natl. Acad. Sci. USA*. 88, 11120-11123.

Kaplan, K.B., Swedlow, J.R., Morgan, D.O., Varmus, H.E. (1995). C-Src enhances the spreading of *src*-/- fibroblasts on fibronectin by a kinase-independent mechanism. *Genes & Developments*. 9, 1505-1517.

Kato, A., Ozawa, F., Saitoh, Y., Hirai, K., Inokuchi, K. (1997). Vesl, a gene encoding VASP/Ena family related protein, is upregulated during seizure, long-term potentiation and synaptogenesis. *FEBS Lett*. 412, 183-189.

Kavanaugh, W.M, Williams, L.T. (1994). An alternative to SH2 domains for binding tyrosine-phosphorylated proteins. *Science*. 266, 1862-1865.

Kay, B.K., Adey, N.B., He, Y.-S., Manfredi, J.P., Mataragnon, A.H., Fowlkes, D.M. (1993). An M13 library displayed 38-amino-acid peptides as a source of novel sequences with affinity to selected targets. *Gene*. 128, 59-65.

Kishan, K.V., Scita, G., Wong, W.T., Di Fiore, P.P., Newcomer, M.E. (1997). The SH3 domain of Eps8 exists as a novel intertwined dimer. *Nature structural biology*. 4 (9), 739-743.

Kitamura, T., Kitamura, Y., Yonezawa, K., Totty, N.F., Gout, I., Hara, K., Waterfield, M.D., Sakaue, M., Ogawa, W., Kasuga, M. (1996). Molecular cloning of p125Nap1, a protein that associates with an SH3 domain of Nck. *Biochem. Biophys. Res. Commun*. 219, 509-514.

Kleinknecht, C., Habib, R. (1992). Nephronophthisis. In *Textbook of clinical nephrology* (Cameron, J.C., Davison, A.M., Grünfeld, J.P., Kers, K.N.S., Ritz, E. Eds.), pp 2188-2197, Oxford Univ. Press, Oxford, England.

Koch Michelsen, B. (1995). Transformation of *Escherichia coli* increases 260-fold upon inactivation of T4 DNA ligase. *Analytical Biochemistry*. 255, 172-174.

Kolanus, W. (1999). The two hybrid toolbox. In *Current Topics in Microbiology and Immunology, 243: Combinatorial Chemistry in Biology*. Famulok, M., Winnacker, E.-L., Wong, C.-H. (Eds.). pp 37-54.

Kolson, D.L., Buchhalter, J. Collman, R., Hellmig, B., Farrell C.F., Debouck, C., Gonzales-Scarano, F. (1993). HIV-1 Tat alters normal organization of neurons and astrocytes in primary rodent brain cell cultures: RGD sequence dependence. *AIDS Res. Hum. Retroviruses*. 9, 677-685.

Kramer, A. Volkmer-Engert, R., Malin, R. Reineke, U., Schneider-Mergener, J. (1993). Simultaneous synthesis of peptide libraries on single resin and continuous membrane supports: identification of protein, metal and DNA binding peptide mixtures. *Peptide Res*. 6, 314-319.

Kramer, A., Schneider-Mergener, J. (1995). Highly complex combinatorial cellulose-bound peptide libraries for the detection of antibody epitopes. In Maia HLS (ed) *Peptides 1994*, Proceedings of the Twenty-Third European Peptide Symposium. ESCOM, Leiden, pp 475-476.

Kremser, A., Rashed, I. (1994). The adsorption of filamentous phage fd: assignment of its disulfide bridges and identification of the domain incorporated in the coat. *Biochemistry*. 33, 13954-13958.

Krook, M., Mosbach, , K., Lindbladh, C., (1994). Selection of peptides with affinity for single-stranded DNA using a phage display library. *Biochem. Biophys. Res. Commun*. 204, 849-854.

Krykbaev, R., McKeating, J., Jones, I. (1997). Mutant CD4 molecules with improved binding to HIV envelope protein gp120 selected by phage display. *Virology*. 234, 196-202.

Kurakin, A., Hoffman, N.G., Kay, B.K. (1998). Molecular recognition properties of the C-terminal SH3 domain of the Cbl associated protein, Cap. *J. Peptide Res*. 52, 331-337.

Ladner, R.C. (1995). Constrained peptides as binding entities. *Tibtech*. 13, 426-430.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of the bacteriophage T4. *Nature*. 227, 680-685.

Lafrenie, R.M, Wahl, L.M:, Epstein, J.S., Hewlett, I.K., Yamada, K.M., Dhawan, S. (1996). HIV-1 Tat modulates the function of monocytes and alters their interactions with microvessel endothelial cells. A mechanism of HIV pathogenesis. *J. Immunol*. 156, 1638-1645.

Lanier, L.M., Gates, M.A., Witke, W., Menzies, A.S., Wehman, A.M., Macklis, J.D., Kwiatkowski, D., Soriano, P., Gertler, F.B. (1999). Mena is required for neurulation and comissure formation. *Neuron*. 22(2), 313-25.

Laurent, V., Loisel, T.P., Harbeck, B., Wehman, A., Gröbe, L., Jockusch, B.M., Wehland, J., Gertler, F.B., Carlier, M.-F. (1999). Role of proteins of the Ena/VASP family in actin-based motility *of Listeria monocytogenes*. *J. Cell Biology*. 144 (6), 1245-1258.

La Vallie, E.R., Rehemtulla, A., Racie, L.A., Diblasio, E.A., Ferenz, C., Grant, K.L., Light, A., McCoy, J.M. (1993) Cloning and functional expression of a cDNA encoding the catalytic subunit of bovine enterokinase. *J. Biol. Chem*. 268, 23311-23317.

Leamon, C.P., Low, P.S. (1992). Cytotoxicity of momordin-folate conjugates in cultured human cells. *J. Biol. Chem.* 267 (35), 24966-71.

Lee, C.-H., Leung, B., Lemmon, M.A., Zheng, J., Cowburn, D., Kuriyan, J., Saksela, K. (1995). A single amino acid in the SH3 domain of Hck determines its high affinity and specificity in binding to HIV-1 Nef protein. *EMBO J*. 14 (20), 5006-5015.

Lee, C.-H., Saksela, K., Mirza, U.A., Chait, B.T., Kuriyan, J. (1996). Crystal structure of the conserved core of HIV-1 Nef complexed with a Src family SH3 domain. *Cell*. 85, 931-942.

Lehmann, J.M., Reithmuller, G., Johnson, J.P. (1990). Nck, a melanoma cDNA encoding a cytoplasmic protein consisting of the src homology units SH2 and SH3. *Nucleic Acids Res*. 18, 1048.

Leung, D.W., Chien, E., Goeddel, D.V. (1989). A method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction. *Technique*. 1 (1), 11-15.

Li, C.J., Friedman, D.J., Wang, C., Metelev, V., Pardee, A.B. (1995). Induction of apoptosis in uninfected lymphocytes by HIV-1 Tat protein. *Science*. 268, 429-431.

Lim, W.A., Richards, F.M., Fox, R.O. (1994). Structural determinants of peptide-binding orientation and of sequence specificity in SH3 domains. *Nature*. 372, 375-379.

Lin, Y.Z., Yao, S.Y., Veach, R.A., Torgerson, T.R., Hawiger, J. (1995). Inhibition of nuclear translocation of transcription factor NF-kappa B by a synthetic peptide containing a cell membrane-permeable motif and nuclear localization sequence. *J. Biol. Chem*. 270, 14255-58.

Ling M.M., Robinson, B.H. (1997). Approaches to DNA mutagenesis: an overview. *Analytical Biochemistry*. 254, 157-178.

Low, N.M., Holliger, P., Winter, G. (1996). Mimicking somatic hypermutation: affinity maturation of antibodies displayed on bacteriophage using a bacterial mutator strain. *J. Mol. Biol*. 260, 359-368.

Lowell, C.A., Soriano, P. (1996). Knockouts of Src-family kinases: Stiff bones, wimpy T cells, and bad memories. *Genes and development*. 10, 1845-1857.

Lowman, H.B., Bass, S.H., Simpson, N, Wells, J.A. (1991). Selecting high-affinity binding proteins by monovalent phage display. *Biochemistry*. 30, 10832-10838.

Lowman, H.B., Wells, J.A. (1991). Monovalent phage display: a method for selecting variant proteins from random libraries. *Methods: a Companion to Methods in Enzymology*. 3, 205-216.

Lowman, H.B., Wells, J.A. (1993). Affinity maturation of human growth hormone by monovalent phage display. *J. Mol. Biol*. 234, 564-578.

Lu, W:, Katz, S., Gupta, R., Mayer, B.J. (1997). Activation of Pak by membrane localization mediated by an SH3 domain from the adaptor protein Nck. *Curr. Biol*. 7, 85-94.

Lussier, G., Larose, L. (1997). A casein kinase I activity is constitutively associated with Nck. *J. Biol. Chem*. 272, 2688-2694.

Ma, J., Ptashne, M. (1987). A new class of transcriptional activators. *Cell.* 51, 113-119.

Maenaka, K., Furuta, M., Tsumoto, K., Watanabe, K., Ueda, Y., Kumagai, I. (1996). A stable phage-display system using a phagemid vector: phage display of hen egg-white lysozyme (HEL), Escherichia coli alkaline phosphatase, and anti-HEL monoclonal antibody, HyHEL10. *Biochem. Biophys. Res. Commun*. 218, 682-687.

Mahoney, N.M., Rozwarski, D.A., Fedorov, E., Fedorov, A.A., Almo, S.C. (1999). Profilin binds proline rich ligands in two distinct amide backbone orientations. *Nature structural biology.* 6, 666-671.

Malin, R., Steinbrecher, A., Semmler, W., Noll, B., Johannsen, B., Frömmel, C., Höhne, W., Schneider-Mergener, J. (1995). Identification of technetium-99 binding peptides using cellulose-bound combinatorial peptide libraries. *Pept. Res*. 9, 151-155.

Malmborg, A.C., Duenas, M., Ohlin, M., Söderlind, E., Borrebaeck, C.A.K. (1996). Selection of binders from phage displayed antibody libraries using BIAcore biosensor. *J. Immunol. Methods*. 198, 51-57.

Malmqvist, M. (1993). Biospecific interaction analysis using biosensor technology. *Nature* 361, 186-187.

Marks, J.D., Griffiths, A.D., Malmqvist, M., Clakson, T.P., Bye, J.M., Winter, G. (1992). By passing immunization: building high affinity human antibodies by chain shuffling. *Bio*/*Technology*. 10, 779-783.

Marston, A.,O. (1986). The purification of eukaryotic polypeptides synthesized in *Escherichia coli*. *Biochem. J*. 240, 1-12.

Mattheakis, L.C., Bhatt, R.R., Dower, W.J. (1994). An *in vitro* polysome display system for identifying ligands from very large peptide libraries. *Proc. Natl. Acad. Sci. USA.* 91, 9022-9026.

Matthews, D.J., Wells, J.A. (1993). Substrate phage: selection of protease substrates by monovalent phage display. *Science*. 260, 1113-1117.

Matuoka, K., Miki, H., Takahashi, K., Takenawa, T. (1997). A novel ligand for an SH3 domain of the adaptor protein Nck bears an SH2 domain and nuclear signaling motifs. *Biochem. Biophys. Res. Commun.* 239, 488-492.

Mayer, B.J., Hirai, H., Sakai, R. (1995).Evidence that SH2 domains promote processive phosphorylation by protein-tyrosine kinases. *Curr. Biol*. 5, 296-305.

McCarty, J.H. (1998). The Nck SH2/SH3 adaptor protein: a regulator of multiple intracellular signal transduction events. *BioEssays*. 20, 913-921.

McConnell, S., Uveges, A., Spinella, D. (1995). Comparison of plate versus liquid amplification of M13 phage display libraries. *Biotechniques*. 18, 803-804.

Medhioub, M., Cherif, D., Benessy, F., Silberman, F., Gubler, M.C., Lepaslier, D., Cohen, D. Weissenbach, J., Beckmann, J., Antignac, C. (1994). Refined mapping of a gene (NPH1) causing familial juvenile nephronophthisis and evidence for genetic heterogeneity. *Genomics*. 22, 296-301.

Meindl, A., Dry, K., Herrmann, K., Manson, F., Ciccodicola, A., Edgar, A., Carvalho, M.R., Achatz, H., Hellebrand, H., Lennon, A., Migliaccio, C., Porter, K., Zrenner, E., Bird, A., Jay, M., Lorenz, B., Wittwer, B., D'Urso, M., Meitinger, T., Wright, A. (1996). A gene (RPGR) with homology to the RCC1 guanine nucleotide exchange factor is mutated in X-linked retinitis pigmentosa (RP3). *Nature Genet*. 13, 35-42.

Meola, A., Delmastro. P., Monaci. P., Luzzago. A., Nicosia. A., Felici. F., Cortese. R., Galfre. G. (1995). Derivation of vaccines from mimotopes. Immunologic properties of human hepatitis B virus surface antigen mimotopes displayed on filamentous phage. *J. Immunol*. 154, 3162-3172.

Minegishi, M., Tachibana, K., Sato, T., Iwata, S., Nojima, Y., Morimoto, C. (1996). Structure and function of Cas-L, a 105-kD Crk-associated substrate-related protein that is involved in beta 1 integrin-mediated signaling in lymphocytes. *J. Exp. Med*. 184, 1365-1375.

Mitchison, T.J., Cramer, L.P. (1996). Actin-based cell motility and cell locomotion. *Cell*. 84, 371-379.

Mitelman, F., Kaneko, Y., Trent, J. (1991). Report of the committee on chromosome changes in neoplasia. Cytogenet. *Cell Genet*. 58, 1053-1079.

Mongiovi, A.M., Romano, P.R., Panni, S., Mendoza, M., Wong, W.T., Musacchio, A,, Cesareni, G., Di Fiore, P.P. (1999). A novel peptide-SH3 interaction. *EMBO J*. 18 (19), 5300-5309.

Model, P., Russel, M. (1988). Filamentous bacteriophage. In *The bacteriophages* (R. Calendar Eds.). Plenum, New York. Vol 2, 375-456.

Mukhija, S., Germeroth, L. Schneider-Mergener, J., Erni, B. (1998). Identification of enzyme I inhibitors of the bacterial phototransferase system using combinatorial cellulose-bound peptide libraries. *Eur. J. Biochem*. 254, 433-438.

Nagahara, H., Vocero-Akbani, A.M., Snyder, E.L., Ho, A., Latham, D.G., Lissy, N.A., Becker-Hapak, M., Ezhevsky, S.A., Dowdy, S.F. (1998). Transduction of full-length TAT fusion proteins into mammalian cells: TAT-p27^{kip1} induces cell migration. *Nature medicine*. 4 (12), 1449-1452.

Nemoto, N., Miyamoto-Sato, E., Husimi, Y., Yanagawa, H. (1997). *In vitro* virus: bonding of mRNA bearing puromycin at the 3´ terminal end to the C-terminal end of its encoded protein on the ribosome *in vitro*. *FEBS Lett*. 414 (2), 405-408.

Nguyen, J.T., Turck, C.W., Cohen, F.E., Zuckermann, R.N., Lim, W.A. (1998). Exploiting the basis of proline recognition by SH3 and WW domains: design of N-substituted inhibitors. *Science*. 282, 2088-2092.

Niebuhr K., Ebel F., Frank R., Reinhard R., Domann E., Carl U.D. Walter U., Gertler F.B., Wehland J., Chakraborty T. (1997). A novel proline-rich motif present in ActA of *Listeria monocytogenes* and cytoskeletal proteins is the ligand for the EVH1 domain, a protein module present in the Ena/VASP family. *EMBO J*. 16, 5433-5444.

Ochs, H.D., Slichter, S.J., Harker, L.A., Von, B.W., Clark, R.A., Wedgwood, R.J. (1980). The Wiskott-Aldrich syndrome: studies of lymphocytes, granulocytes, and platelets. *Blood*. 55, 243-252.

Ohlin, M., Owman, H., Mach, M., Borrebaeck, C.A.K. (1996). Light chain shuffling of a high affinity antibody results in a drift in epitope recognition. *Molecular Immunology*. 33 (1), 47-56.

Oligino, L., Lung, F.-D.,T., Sastry, L., Bigelow, J., Cao, T., Curran, M., Burke, T.R., Jr., Wang, S., Krag, D., Roller, P.P., King, C.R. (1997). Nonphosphorylated peptide ligands for the Grb2 Src homology 2 domain. *J. Biol. Chem*. 272, 29046-29052.

Panayotou, G., Waterfield, M.D., End, P. (1993). Riding the evanescent wave. *Current Biology*. 3 (12), 913-915.

Pannekoek, H., Van Meijer, M., Schleef, R.R., Loskutoff, D.J., Barbas, C. F. III (1993). Functional display of human plasminogen-activator inhibitor 1 (PAI-1) on phages: novel perspectives for structure-function analysis by error-prone DNA synthesis. *Gene*. 128, 135-140.

Pasqualini, R., Koivunen, E., Ruoslahti, E. (1995). A peptide isolated from phage display libraries is a structural and functional mimic of an RGD-binding site on integrins. *J. Cell. Biol.* 130, 1189-1196.

Pawson, T. (1995). Protein modules and signaling networks. *Nature*. 373, 573-580.

Pawson, T., Gish, G.D. (1992). SH2 and SH3 domains: from structure to function. *Cell*. 71, 359-362.

Pendergast, A.M., Quilliam, L.A., Cripe, L.D., Bassing, C., H., Dai, Z., Li, N., Batzer, A., Rabun, K.M., Der, C.J., Schlessinger, J., Gishizky, M.L. (1993). Bcr-Abl-induced oncogenesis is mediated by direct interaction with the SH2 domain of the Grb2 adaptor protein. *Cell*. 75, 175-185.

Perlmutter, R.M., Alberola-Ila, J. (1996). The use of dominant-negative mutations to elucidate signal transduction pathways in lymphocytes. *Curr. Opin. Immunol*. 8, 285-290.

Pierce, H.H., Adey, N., Kay, B.K. (1996). Identification of cyclized calmodulin antagonists from a phage display random peptide library. *Mol. Divers*. 1, 259-265.

Pisabarro, M.T., Serrano, L. (1996). Rational design of specific high-affinity peptide ligands for the Abl-SH3 domains. *Biochemistry*. 35, 10634-10640.

Powers, M.A., Forbes, D.J. (1994). Cytosolic factors in nuclear transport: what's Importin? *Cell*. 79, 931-934.

Prehoda, K.E., Lee, D.J., Lim, W.A. (1999). Structure of the Enabled/VASP homology 1 domain-peptide complex: a key component in the spatial control of actin assembly. *Cell*. 97, 471-480.

Prior, T.I., Fitzgerald, D.J., Pastan, I. (1992). Translocation mediated by domain II of Pseudomonas exotoxin A: transport of barnase into the cytosol. *Biochemistry*. 31, 3555-3559.

Puil, L., Liu, J., Gish, G., Mbamalu, G:, Bowtell, D., Pelicci, C.G., Arlinghaus, R., Pawson, T. (1994). Bcr-Abl oncoproteins bind directly to activators of the Ras signalling pathway. *EMBO J.* 13, 764-773.

Quilliams, L.A., Lambert, Q.T., Mickelson-Young, L.A., Westwick, J.K., Sparks, A.B., Kay, B.K., Jenkins, N.A., Gilbert, D.J., Copeland, N.G., Derr, C.J. (1996). Isolation of a Nck-associated kinase, PRK2, an SH3-binding protein and potential effector of Rho protein signaling. *J. Biol. Chem*. 271, 28772-28776.

Reineke, U., Sabat, R., Volk, H.D., Schneider-Mergener, J.(1998). Mapping of the interleukin-10/interleukin-10 receptor combining site. *Prot. Sci*. 7, 951-960.

Reineke, U., Kramer, A., Schneider-Mergener, J. (1999). Antigen sequence- and library-based mapping of linear and discontinuous protein-protein interaction sites by spot synthesis. In *Current Topics in Microbiology and Immunology, 243: Combinatorial Chemistry in Biology*. Famulok, M., Winnacker, E.-L., Wong, C.-H. (Eds). pp 24-36.

Reinhard, M., Halbrugge, M:, Scheer, U., Wiegand, C., Jockusch, B.M., Walter, U. (1992). The 46/50 kDa phosphoprotein VASP purified from human platelets is a novel protein associated with actin filaments and focal contacts. *EMBO. J.* 11, 2063-2070.

Reinhard, M., Giehl, K., Abel, K., Haffner, C., Jarchau, T., Hoppe, V., Jockusch, B.M., Walter, U. (1995a). The proline-rich focal adhesion and microfilament protein VASP is a ligand for profilins. *EMBO. J.* 14, 1583-1589.

Reinhard, M., Jouvenal, K., Tripier, D., Walter, U. (1995b). Identification, purification, and characterization of a zyxin-related protein that binds the focal adhesion and microfilament VASP (vasodilator-stimulated phosphoprotein). *Proc. Natl. Acad. Sci. USA.* 92, 7956-7960.

Reinhard, M., Rudiger, M., Jockusch, B.M., Walter, U. (1996). VASP interaction with vinculin: a recurring theme of interactions with proline-rich motifs. *FEBS Lett.* 399, 103-107.

Ren, R., Mayer, B.J., Cicchetti, P., Baltimore, D. (1993). Identification of a ten-amino acid proline rich SH3 binding site. *Science.* 259, 1157-1161.

Renneisen, K., Leserman, L., Matthes, E., Schroder, H.C., Muller, W.E. (1990). Inhibition of expression of human immunodeficiency virus-1 in vitro by antibody-targeted liposomes containing antisense RNA to the env region. *J. Biol. Chem.* 265, 16337-16342.

Reuter, M., Schneider-Mergener, J., Kupper, D., Meisel, A., Mackeldanz, P., Krüger, D., Schroeder, C. (1999). EcoRII DNA target recognition sites identified by membrane-bound peptide repertoires. *J. Biol. Chem.* 274, 5213-5221.

Rickles, R.J., Botfield, M.C., Weng, Z., Taylor, J.A., Green, O.M., Brugge, J.S., Zoller, M.J. (1994). Identification of Src, Fyn, Lyn, PI3K, and Abl SH3 domain ligands using phage display libraries. *EMBO J.* 13, 5598-5604.

Rickles, R.J., Botfield, M.C., Zhou, X.M., Henry, P.A., Brugge, J.S., Zoller, M.J. (1995). Phage display selection of ligand residues important for Src Homology 3 domain binding specificity. *Proc. Natl. Acad. Sci. USA.* 92, 10909-10913.

Rivero-Lezcano, O.M., Sameshima, J.H., Marcilla, A., Robbins, K.C. (1994). Physical association between SH3 elements and the protein product of the c-cbl proto-oncogene. *J. Biol. Chem*. 269, 17363-17366.

Rivero-Lezcano, O.M., Marcilla, A., Sameshima, J.H., Robbins, K.C. (1995). Wiscott-Aldrich syndrome protein physically associates with Nck through SH3 domains. *Mol. Cell. Biol*. 15, 5725-5731.

Roberts, B.L. Markland. W., Ley. A.C., Kent. R.B., White. D.W., Guterman. S.K., Ladner. R.C. (1992). Directed evolution of a protein: selection of potent neutrophil elastase inhibitors displayed on M13 fusion phage. *Proc. Natl. Acad. Sci. USA*. 89, 2429-2433.

Roberts, R.W., Szostak, J.W. (1997). RNA-peptide fusions for the *in vitro* selection of peptides and proteins. *Proc. Natl. Acad. Sci*. 94 (23), 12297-12302.

Rojas, M., Donahue, J.P., Tan, Z., Lin, Y.-Z. (1998). Genetic engineering of proteins with cell membrane permeability. *Nature Biotechnology*. 16, 370-375.

Röttgen, P., Collins, J. (1995). A human pancreatic secretory trypsin inhibitor presenting a hypervariable highly constrained epitope via monovalent phagemid display. *Gene*. 164, 243-250.

Röttgen, P. (1996). Phagenoberflächenpräsentation und Affinitätsselektion hypermutierter Varianten des humanen pankreatischen sekretorischen Trypsin-Inhibitors. PhD thesis, Westfälische Willhelms-Universität Münster.

Rüdiger, S., Germeroth, L., Schneider-Mergener, J., Bukau, B. (1997). Substrate specificity of the Dank chaperone determined by screening of cellulose-bound peptide libraries. *EMBO J*. 16, 1501-1507.

Saksela, K., Cheng, G., Baltimore, D. (1995). Proline rich (PXXP) motifs in HIV-1 Nef bind to SH3 domains of a subset of Src kinases and are required for the enhanced growth of Nef+ viruses but not for down-regulation of CD4. *EMBO J*. 14 (3), 484-491.

Sambrook, J., Fritsch, E.F., Maniatis, T. (1989). *In vitro* amplification of DNA by the polymerase chain reaction. In *Molecular cloning: a laboratory manual*, second edition, pp 14.2-14.35. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Schein, C.H. (1989) Production of soluble recombinant proteins in bacteria. *Bio*/*Technology*. 7, 1141-1149.

Schumacher, T.N.M., Mayr, L.M., Minor, D.L., Jr., Milhollen, M.A., Burgess, M.W., Kim, P.S. (1996). Identification of D-peptide ligands through mirror image phage display. *Science*. 271, 1854-1857.

Schwartzberg, P.L., Xing, L., Hoffmann, O., Lowell, C.A., Garrett, L., Boyce, B.F., Varmus, H.E. (1997). Rescue of osteoclast function by transgenic expression of kinase-deficient Src in *src*-/- mutant mice. *Genes & Developments*. 11, 2835-2844.

Schwarze, S.R., Ho, A., Vocero-Akbani, A., Dowdy, S.F. (1999). *In vivo* protein transduction: delivery of a biologically active protein into the mouse. *Science*. 285, 1569-1572.

She, H., Rockow, S., Tang, J., Nishimura, R., Skolnik, E.Y., Chen, M., Margolis, B., Lei, W. (1997). Wiskott-Aldrich syndrome protein is associated with the adapter protein Grb2 and the epidermal growth factor in living cells. *Mol. Cell. Biol*. 8, 1709-1721.

Shima, F., Yamawaki-Kataoka, Y., Yanagihara, C., Tamada, M., Okada, T., Kariya, K.-I., Kataoka, T. (1997). Effect of association with adenylyl cyclase-associated protein on the interaction of yeast adenylyl cyclase with Ras protein. *Mol. Cell. Biol*. 17, 1057-1064.

Short, J.M., Fernandez, J.M., Sorge, J.A., Huse, W.D. (1988). Lambda ZAP: a bacteriophage lambda expression vector with in vivo excision properties. *Nucleic Acids Res*. 16, 7583-7600.

Sibille, P.,Ternynck, T., Nato, F., Buttin, G., Strosberg, D., Avrameas, A. (1997). Mimotopes of polyreactive anti-DNA antibodies identified using phage-display peptide libraries. *Eur. J. Immunol.* 27, 1221-1228.

Simons, P.C. and Vander Jagt, D.L. (1977) Purification of glutathione S-transferases from human liver by glutathione-affinity chromatography. *Anal. Biochem*. 82, 334-341.

Smith, G.P (1985). Filamentous fusion phage: Novel expression vectors that display cloned antigens on the surface of the virion. *Science*. 228, 1315-1317.

Smith, G.P. (1993). Surface display and peptide libraries. *Gene*. 128, 1-2.

Smith, D.B., Johnson, K.S. (1988). Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. *Gene*. 67, 31-40.

Soriano, P., Montgomery, C., Geske, R., Bradley, A. (1991). Targeted disruption of the *c-src* proto-oncogene leads to osteopetrosis in mice. *Cell*. 64, 693-702.

Sparks, A.B. Quilliam, L.A., Thorn, J.M., Der, C.J., Kay, B.K. (1994). Identification and characterization of Src SH3 ligands from phage-displayed random peptide libraries. *J. Biol. Chem*. 269, 23853-23856.

Sparks, A.B., Rider, J.E., Hoffman, N.G., Fowlkes, D.M., Quilliam, L.A., Kay, B.K. (1996). Distinct ligand preferences of Src homology 3 domains from Src, Yes, Cortactin, p53bp2, PLCγ, Crk, and Grb2. *Biochemistry* 93, 1540-1544.

Stanssens, P., Opsomer, C., McKeown, Y.M., Kramer, W., Zabeau, M., Fritz, H.-J. (1989). Efficient oligonucleotide-directed construction of mutations in expression vectors by the gapped duplex DNA method using alternating selectable markers. *Nucl. Acids Res.* 17, 4441-4448.

Stauffer, T.P., Martenson, C.H., Rider, J.E., Kay, B.K., Meyer, T. (1997). Inhibition of Lyn function in mast cell activation by SH3 domain binding peptides. *Biochemistry*. 36, 9388-9394.

Stemmer, W.P.C. (1994a). Rapid evolution of a protein in vitro by DNA shuffling. *Nature.* 370, 389-391.

Stemmer, W.P.C. (1994b). DNA shuffling by random fragmentation and reassembly: *in vitro* recombination for molecular evolution. *Proc. Natl. Acad. Sci. USA*. 91, 10747-10751.

Stenmark, H., Moskaug, J.O, Madshus, I.H., Sandvig, K., Olsnes, S. (1991). Peptides fused to the amino-terminal end of diphtheria toxin are translocated to the cytosol. *J.Cell. Biol*. 113, 1025-1032.

Stormo, G.D., Schneider, T.D., Gold, L. (1982) Characterization of translation initiation sites in *E. coli*. *Nucl. Acids Res.* 10, 2971-2996.

Studier, F.W., Rosenberg. A.H., Dunn. J.J., Dubendorff. J.W. (1990). Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol*. 185, 60-89.

Su, Y.-C., Han, J., Xu, S., Cobb, M., Skolnik, E.Y. (1997). NIK is a new Ste20-related kinase that binds Nck and Mekk1 and activate the Sapk/Jnk cascade via a conserved regulatory mechanism. *EMBO J.* 16, 1279-1290.

Superti-Furga, G., Fumagalli, S., Koegl, M., Courtneidge, S.A., Draetta, G. (1993). Csk inhibition of c-Src activity requires both the SH2 and SH3 domains of Src. *EMBO J*. 12, 2625-2634.

Symons, M., Derry, J.M.J., Karlak, B., Jiang, S., Lemahieu, V., McCormick, F., Francke, U., Obo, A. (1996). Wiskott-Aldrich syndrome protein, a novel effector for the GTPase cdc42hs, is implicated in actin polymerisation. *Cell*. 84, 723-734.

Szardenings, M., Collins, J. (1990). A phasmid optimized for protein design projects: pMAMPF. *Gene*. 94, 45-52.

Tegge, W.J., Frank, R. (1998). Analysis of protein kinase substrate specificity by the use of peptide libraries on cellulose paper (SPOT-method). *Meth. Mol. Biol*. 87, 99-106.

Tesar, M., Beckmann, C., Röttgen, P., Haase, B., Faude, U., Timmis, K.N. (1995). Monoclonal antibody against pIII of filamentous phage: an immunological tool to study pIII fusion protein expression in phage display systems. *Immunotechnology*. 1, 53-64.

Tokunaga, K., Ikuta, K., Adachi, A., Matsuda, M., Kurata, T., Kojima, A. (1999). The cellular kinase binding motifs (PxxP andRR) in human immunodeficiency virus type 1 Nef protein are dispensable for producer-cell-dependent enhancement of vira entry. *Virology*. 257, 285-289.

Toomik, R., Edlund, M., Ek, P., Obrink, B., Engstrom, L. (1996). Simultaneously synthesised peptides on continuous cellulose membranes as substrates for protein kinases. *Pept. Res*. 9, 6-11.

Tsunetsugu-Yokota, Y., Tatsumi, M., Robert, V., Devaus, C., Sprite, B., Chermann, J.-C., Hirsch, I. (1991). Expression of an immunogenic region of HIV by a filamentous bacteriophage vector. *Gene*. 99, 323-326.

Tsurushita, M., Fu, H., Warren, C. (1996). Phage display vectors for *in vivo* recombination of immunoglobulin heavy and light chain genes to make large combinatorial libraries. *Gene*. 172, 59-63.

Tu, J.C., Xiao, B., Yuan, J.P., Lanahan, A.A., Leoffert, K., Li, M., Linden, D.J., Worley, P.F. (1998). Homer binds a novel proline-rich motif and links group 1 metabotropic glutamate receptors with IP3 receptors. *Neuron*. 21, 717-726.

Tu, J.C., Xiao, B., Naisbitt, S., Yuan, J.P., Petralia, R.S., Brakeman, P., Doan, A., Aakalu, V.K., Lanahan, A.A., Sheng, M., Worley, P.F. (1999). Coupling of mGluR/Homer and PSD-95 complexes by the Shank family of postsynaptic density proteins. *Neuron*. 23, 583-592.

Van Wezenbeek, P.M.G.F., Hulsebos, T.J.M., Schoenmakers, J.G.G. (1980). Nucleotide sequence of the filamentous bacteriophage M13 DNA genome: Comparison with phage fd. *Gene*. 11, 129-148.

Viera, J., Messing, J. (1987). Production of single-stranded plasmid DNA. *Meth. Enz*. 153, 3-11.

Viguera, A.R., Arrondo, J.L.R., Musacchio, A., Saraste, M., Serrano, L. (1994). Characterization of the interaction of natural proline-rich peptides with five different SH3 domains. *Biochemisty*. 33, 10925-10933.

Vivès, E., Brodin, P., Lebleu, B. (1997). A truncated HIV-1 Tat protein basic domain rapidly translocates through the plasma membrane and accumulates in the cell nucleus. *J. biol. Chem*. 272 (25), 16010-16017.

Vocero-Akbani, A.M., Vander Heyden, N., Lissy, N.A., Ratner, L., Dowdy, S.F. (1999). Killing HIV-infected cells by transduction with an HIV protease-activated caspase-3 protein. *Nature Medicine*. 5, 29-33.

Wang, K., Feramisco, J.R., Ash, J.F. (1982). Fluorescent localization of contractile proteins in tissue culture cells. *Methods Enzymol*. 85, 514-562.

Wang, L.-F., Du Plessis, D.H., White, J.R., Hyatt, A.D., Eaton, B.T. (1995). Use of a gene-targeted phage display random epitope library to map an antigenic determinant on the bluetongue virus outer capsid protein VP5. *J. Immunol. Methods*. 178, 1-12.

Waterhouse, P., Griffiths, A.D., Jonhson, K.S., Winter, G. (1993). Combinatorial infection and *in vivo* recombination: a strategy for making large phage antibody repertoires. *Nucleic Acids Res*. 21, 2265-2266.

Webster, R.E. (1996). Biology of the filamentous bacteriophage. In *Phage display of peptides and proteins: a laboratory manual*. Kay, B.K., Winter J., McCafferty, J. Eds. Academic press, Inc, San Diego. 1-20.

Wilson, J.M., Grossman, M., Wu, C.H., Chowdhury, N.R., Wu, G.Y., Chowdhury, J.R. (1992). Hepatocyte-directed gene transfer in vivo leads to transient improvement of hypercholesterolemia in low density lipoprotein receptor-deficient rabbits. *J. Biol. Chem*. 267, 963-967.

Wrighton, N.C., Farrell, F.X., Chang, R., Kashyap, A.K., Barbone, F.P., Mulcahy, L.S., Johnson, D.L., Barrett, R.W., Jolliffe, L.K., Dower, W.J. (1996). Small peptides as potent mimetics of the protein hormone erythropoietin. *Science*. 273, 458-464.

Wu, G.Y., Wu, C.H. (1991). Delivery systems for gene therapy. *Biotherapy*. 3, 87-95.

Wunderlich, L., Farago, A., Buday, L. (1999a). Characterization of interactions of Nck with Sos and Dynamin. *Cell. Signal*. 11 (1), 25-29.

Wunderlich, L., Goher, A., Farago, A., Downward, J., Buday, L. (1999b). Requirement of multiple SH3 domains of Nck for ligand binding. *Cell Signal*. 11 (4), 253-262.

Xu, W., Harrison, S.C., Eck, M.J. (1997). Three dimensional structure of the tyrosine kinase c-Src. *Nature*. 385, 595-602.

Yanish-Perron, C., Viera, J., Messing, J. (1985). *Gene*. 33, 103-199.

Yanofsky, S.D., Baldwin, D.N., Butler, J.H., Holden, F.R., Jacobs, J.W., Balasubramanian, P., Chinn, J.P., Cwirla, S.E:, Peters-Bhatt, E., Whitehorn, E,A,, Tate, E.H., Akeson, A., Bowlin, T.L., Dower, W.J. (1996). High affinity type I interleukin 1 receptor antagonists discovered by screening recombinant peptide libraries. *Proc. Natl. Acad. Sci. USA*. 93, 7381-7386.

Ymer, S. (1991). Heat inactivation of DNA ligase prior to electroporation increases transformation efficiency. *Nucl. Acids Res*. 19, 6960.

Yu, H., Chen, J.K., Feng, S., Dalgarno, D.C., Brauer, A.W., Schreiber, S.L. (1994). Structural basis for the binding of proline-rich peptides to SH3 domains. *Cell*. 76, 933-945.

Yu, J., Wang, C., Palmieri, S.J., Haarer, B.K., Field, J. (1999). A cytoskeletal localizing domain in the cyclase-associated protein CAP/Srv2p, regulates access to a distant SH3 binding site. *J. Biol. Chem*. 274 (28), 19985-19991.

Zell, R., Fritz, H.-J. (1987). DNA mismatch repair in *Escherichia coli* counteracting the hydrolytic deamination of 5-methylcystosine residues. *EMBO J.* 6, 1809-1815.

VIII - ABBREVIATIONS

1- General abbreviations

A	Adenosine/Desoxyadenosine (-phosphate)
Amp	Ampicilin
Amp^R	Ampicilin resistance
APS	Ammoniumpersulfate(-solution)
bla	13-Lactamase gene (ampicillin resistance)
bp	Basepairs
BSA	Bovine Serum Albumin
C	Cytidine/Desoxycytidine (-phosphate)
CAP1	Adenylyl Cyclase Associated Protein 1
cat	Gene encoding for chloramphenicol-acetyltransferase
cDNA	Complementary DNA
CDR	Complementary Determining Region
cfu	Colony Forming Unit
CPPI	Cell-Permeable Peptide Import
ddH_2O	Double Deionized Water
ddNTP	Dideoxyribonucleoside triphosphate
dNTP	Deoxyribonucleoside triphosphate
DMEM	Dulbecco's Modified Eagle's Medium
DMSO	Dimethylsulfoxide
DNA	DeoxyriboNucleic Acid
DTT	Dithiothreitol
E. coli	*Escherichia coli*
EDC	N-ethyl-N'-(3-dimethylaminopropyl)carbodiimide hydrochloride
EDTA	Ethylendiamintetraacetate
EGF	Epidermal Growth Factor
EGTA	Ethylenbis(oxyethylennitrilo)tetraacetic acid
ELISA	Enzyme Linked ImmunoSorbent Assay
EVH1	Ena-VASP Homology domain 1
Evl	Ena-VASP-Like
FACS	Fluorescent Activated Cell Sorting

FITC	Fluorescein IsoThioCyante
FGF	Fibroblast Growth Factor
G	Guanosine/Desoxyguanosine (-phosphate)
GAP3DH	Glyceraldehyde-3-phosphate dehydrogenase
GBF	Gesellschaft für Biotechnologische Forschung
GST	Glutathione-S-Transferase
h	Hour
h region	Hydrophobic region
HAc	Acetic acid
HEPES	N-(2-Hydroxyethyl)-piperazine- N'-2-ethane sulfonic acid
HIV-1	Human Immunodeficiency Virus Type 1
HRP	HorseRadish Peroxidase
IP3	Inositol Triphosphate
IPTG	Isopropyl-β-D-thiogalactopyranoside
kb	Kilobase
K_D	Dissociation constant
KDa	Kilodalton
Km^R	Kanamycin resistance
LB	Luria-Bertina medium
L. monocytogenes	*Listeria monocytogenes*
λP_L	Major leftward promoter of the bacteriophage λ
MBP	Maltose-Binding Protein
mGluR	Group 1 Metabotropic Glutamate Receptor
MGR1	Metabotropic Glutamate Receptor 1
min	Minute
MPS	Magnetic Particle Concentrator
MTS	Membrane Translocating Sequence
Nap1	Nck-Associated Protein1
NHS	N-hydroxysuccinimide
NLS	Nuclear Localization Signal
NP1	Clone Non Proline 1
NP2	Clone Non Proline 2
NMR	Nuclear Magnetic Resonance
NPH	Nephronophthisis

NPH1	Gene locus for nephronophthisis
NPHP1	Gene locus for nephronophthisis often deleted in NPH1 families
OD	Optical density
ompL	Sequence of *E. coli* derived from the outer membrane protein OmpA
ori fd	Origin of replication for single stranded DNA of the bacteriophage fd
ori E	Origin of replication for plasmids (colE1)
PAGE	PolyAcrylamide Gel Electrophorese
PBS	Phosphate Buffer Saline
PCR	Polymerase Chain Reaction
PEG	Polyethylene Glycol 6000
pfu	Plaque Forming Unit
PPII	Polyproline Type II helix
PSTI	Human Pancreatic Secretory Trypsin Inhibitor
RF	Replicative Form
RHG1	Rho-GTPase Activating Protein 1
rpm	Rounds Per Minute
scFv	Single Chain Fv antibody fragment
SDS	Sodium Dodecyl Sulfate
sec	Seconde
SH2	Src Homology 2
SH3	Src Homology 3
SPR	Surface Plasmon Resonance
T	Thymidine
TAE	Tris/Sodium acetate/EDTA
Tc^R	Tetracycline resistance
TE	Tris/EDTA
TEMED	N,N,N',N'-Tetramethyl-ethylendiamin
TRAP	Tartrate Resistant Acid Phosphatase
Tris	Tris-(hydroxymethyl)-aminomethane
Trx	Thioredoxine
U	Unit
UV	Ultra Violet
V	Volt
VASP	Vasodilator-Stimulated Phosphoprotein

VEGF	Vascular Endothelial cell Growth Factor
v/v	Volume/volume
WAP	Wiskott-Aldrich Syndrome
WASP	Wiskott-Aldrich Syndrome protein
WH2	WASP Homology 2
WIP	WASP Interacting Protein
w/v	Weight/volume

2- Amino acid codes

A	Ala	Alanine	M	Met	Methionine
C	Cys	Cysteine	N	Asn	Asparagine
D	Asp	Aspartic Acid	P	Pro	Proline
E	Glu	Glutamic Acid	Q	Gln	Glutamine
F	Phe	Phenylalanine	R	Arg	Arginine
G	Gly	Glycine	S	Ser	Serine
H	His	Histidine	T	Thr	Threonine
I	Ile	Isoleucine	V	Val	Valine
K	Lys	Lysine	W	Trp	Tryptophan
L	Leu	Leucine	Y	Tyr	Tyrosine

ACKNOWLEDGMENTS

My first thanks go to Prof. Dr. John Collins (GBF, Cosmix GmbH, Braunschweig) for the opportunity he gave me to work and learn in his laboratory, and to entrust this fascinating project to me.

I would like to give a very special thanks to Dr. Peter Röttgen (DKFZ Heidelberg, formerly Cosmix GmbH) and Dr. Michael Szardenings (Cosmix GmbH, Braunschweig) for their advises and helpful discussions. They were present each time I needed them, and even more.

A particular thanks to the members of my laboratory, Dr. Thomas Böldicke, Doris Stellfeld, and most of all, Walburga Westphal for their presence and encouragement. Thanks to Walli for the technical help she gave me.

Thanks also to Dr. Gerhard Gross (GBF) for his support and for his optimism.

Thanks for kindly providing us various GST-SH3 fusion domains, to Dr. Brian B. Kay and Dr. Lawrence Quilliams (University of Wisconsin-Madison, Wisconsin, USA); to Prof. Dr. Jürgen Wehland and Dr. Uwe Carl (GBF, Braunschweig) for the GST-EVH1 fusion domains; to Dr. Friedhelm Hildebrandt and Dr. Edgar Otto (University of Freiburg) for the nephrocystin SH3 domain cDNA; and to Dr. Steven F. Dowdy and Dr. Michelle Becker-Hapak (University of Washington, Saint Louis, Missouri, USA) for the pTAT-HA vector.

Thanks to Dr. Kurt Dittmar (GBF) for his help with the confocal microscope, and to Dr. Ronald Frank (GBF) for the synthesis of peptides.

I also would like to thank the members of my new laboratory, the group of Dr. Lydia Sorokin, in Erlangen, for the nice atmosphere they bring to the lab, and the energy I could found there.

I want to give a very big thanks to Clemens who always knew how to help me during the work, and knew (or learned) how to handle with my “very special” mood.

And finally, I would like to thanks my parents for their confidence and encouragement all along my studies. Pour votre confiance, mille mercis.

Thanks, Nathalie

CURRICULUM VITAE

Name: **Nathalie HORN** (born Chavand)
Date of birth: 09/09/1972
In: Grenoble/France
Nationality: French
Marital status: married

Education:

1975-1978: Ecole maternelle Barnave - Saint Egrève.
1978-1982: Ecole primaire Barnave - Saint Egrève.
1982-1986: Collège Barnave - Saint Egrève.
1986-1989: Lycée Champollion - Grenoble.
06/1990: Baccalauréat D (option Mathematics and Life Sciences). Cum Laude.

Scientific education: Biology studies

09/1990 - 06/1994: University of Joseph Fourier, Grenoble, France

1991-1992: First year of Medicine.
06/1993: Deug Biologie (equivalent to a 2 year degree in Life sciences). Cum Laude.
06/1994: Licence de Biologie Cellulaire et Physiologie (equivalent to a Bachelor's degree of Cellular Biology and Physiology).

09/1994 - 06/1995: University of Sussex, Brighton, United Kingdom

06/1995: Maitrise de Biologie Cellulaire et Physiologie (equivalent to a Master's degree of Cellular Biology and Physiology). 2/i honours degree.

09/1995 - 02/1996: University of Luminy, Marseille, France
03/1996 - 09/1996: Nestlé, Lausanne, Switzerland

1996: DESS Immunotechnologie. Theoretical part of the studies in Marseille, France.
6 month Research Project in Nestlé Research Center, Lausanne, Switzerland.
Establishment of a T cell line protecting the organism against the βlactoglobulin.
09/1996: DESS Diploma in Immunotechnology (equivalent to a doctoral level study in Immunotechnology). Cum Laude.

10/1996 – 10/1999: Experimental part of the PhD, under the supervision of Prof. Dr. John COLLINS, at GBF, Braunschweig, Germany

Isolation of high affinity ligands against SH3 and EVH1 domains, using the phage display cosmix-plexing® technology.